Evidence Report/Technology Assessment
Number 130

Management of Adnexal Mass

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-02-0025

Prepared by:
Duke Evidence-based Practice Center
Durham, North Carolina

Investigators
Evan R. Myers, MD, MPH
Lori A. Bastian, MD, MPH
Laura J. Havrilesky, MD
Shalini L. Kulasingam, PhD
Mishka S. Terplan, MD
Kathryn E. Cline, MHS
Rebecca N. Gray, DPhil
Douglas C. McCrory, MD, MHS

AHRQ Publication No. 06-E004
February 2006

This report is based on research conducted by the Duke Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-02-0025). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

Suggested Citation:
Myers ER, Bastian LA, Havrilesky LJ, Kulasingam SL, Terplan MS, Cline KE, Gray RN, McCrory DC. Management of Adnexal Mass. Evidence Report/Technology Assessment No. 130 (Prepared by the Duke Evidence-based Practice Center under Contract No. 290-02-0025.) AHRQ Publication No. 06-E004. Rockville, MD: Agency for Healthcare Research and Quality. February 2006.

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-Based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. This report was requested and funded by the Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion at the Centers for Disease Control and Prevention (CDC). The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To bring the broadest range of experts into the development of evidence reports and health technology assessments, AHRQ encourages the EPCs to form partnerships and enter into collaborations with other medical and research organizations. The EPCs work with these partner organizations to ensure that the evidence reports and technology assessments they produce will become building blocks for health care quality improvement projects throughout the Nation. The reports undergo peer review prior to their release.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

We welcome comments on this evidence report. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by e-mail to **epc@ahrq.gov.**

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Eddie Reed, M.D.
Director, Division of Cancer Prevention and Control
National Center for Chronic Disease Prevention and Promotion
Centers for Disease Control and Prevention

Beth A. Collins Sharp, R.N., Ph.D.
Acting Director, EPC Program
Agency for Healthcare Research and Quality

Susan Meikle, M.D., M.S.P.H.
EPC Program Task Order Officer
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge Jane Kolimaga and Alison Lee for assistance in study initiation; Margaret Jamison, PhD, for assistance with analysis of the Nationwide Inpatient Sample data set; and Karen Hoffman, MD, for initial work on developing the ovarian cancer natural history model. They also thank Mona Saraiya, MD, MPH, and Christie Eheman, PhD, from the Centers for Disease Control and Prevention for their valuable input.

Structured Abstract

Objectives: To assess diagnostic strategies for distinguishing benign from malignant adnexal masses.

Data Sources: MEDLINE® and reference lists of recent reviews; discharge data from the Nationwide Inpatient Sample.

Review Methods: The major diagnostic methods evaluated were bimanual pelvic examination, ultrasound (morphology and Doppler velocimetry), MRI, CT, FDG-PET, CA-125, and scoring systems that incorporated multiple clinical, laboratory, and radiologic findings. Meta-analysis using a random-effects model was used to estimate pooled sensitivity and specificity for discriminating benign from malignant. We reviewed evidence for followup strategies for masses considered benign, and for adverse outcomes of diagnostic surgery. We also reviewed published models of the natural history of ovarian cancer and compared the impact of assumptions about natural history on outcomes.

Results: The majority of studies did not describe whether patients presented with asymptomatic masses detected through screening or with symptoms. Prevalence of malignant masses in a U.S. postmenopausal screening population was approximately 0.1 percent, while benign masses were found in 0.8 to 1.8 percent of women. Pooled (a) sensitivity and (b) specificity were: bimanual exam (a) 0.45, (b) 0.90; ultrasound morphology scores (a) 0.86 to 0.91, (b) 0.68 to 0.83; Doppler resistive index (a) 0.72, (b) 0.90; pulsatility index (a) 0.80, (b) 0.73; maximum systolic velocity (a) 0.74, (b) 0.81; presence of vessels (a) 0.88, (b) 0.78; combined morphology and Doppler (a) 0.86, (b) 0.91; MRI (a) 0.91, (b) 0.88; CT (a) 0.90, (b) 0.75; FDG-PET (a) 0.67, (b) 0.79; and CA-125 (a) 0.78, (b) 0.78. Both sensitivity and specificity of CA-125 were better in postmenopausal than in premenopausal women. In modeled outcomes, combinations of imaging and CA-125 were both more sensitive and more specific than either alone. Performance of scoring systems in validation studies was consistently worse than in development studies; the highest demonstrated specificity observed was 0.91, with a concurrent sensitivity of 0.74. Evidence on followup strategies was sparse, although one large study provided good evidence for safely following unilocular cysts less than 10 cm in diameter. Overall complication rates in studies of surgically managed adnexal masses were low, but important clinical information was not reported.

Conclusions: All diagnostic modalities showed trade-offs between sensitivity and specificity, but the available literature does not provide sufficient detail on relevant characteristics of study populations to allow confident estimation of the results of alternative diagnostic strategies. Although modeling studies may prove useful in evaluating diagnostic algorithms, further work is needed to explore the implications of uncertainty about the natural history of ovarian cancer.

Contents

Figures

Tables

Appendixes

The Appendixes and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/adnexal/adnexal.pdf

Executive Summary

Introduction

Ovarian cancer is the leading cause of cancer death from gynecologic malignancies in the United States, with an annual incidence of over 25,000 and an annual mortality of approximately 14,000. Cancer incidence increases dramatically with age.

The high case-fatality rate has largely been attributed to the fact that most ovarian cancers are diagnosed in advanced stages (Stage III, where the cancer has spread beyond the pelvis to organs of the upper abdominal cavity, and Stage IV, where the cancer has spread outside of the peritoneal cavity), when survival is poor. Stage I cancer (limited to the ovaries) has a survival rate of over 90 percent. Thus, there has long been an emphasis on early detection of ovarian cancer in the belief that detection in early stages will lead to decreases in morbidity and mortality. The detection of a mass in the area of the ovaries and fallopian tubes (the uterine adnexae) raises the possibility of ovarian cancer, which necessitates further study to rule out malignancy.

There are two main clinical routes by which an adnexal mass may be detected: (1) women with symptoms may have an adnexal mass detected as part of their evaluation for those symptoms, either by physical exam or radiographic imaging; (2) the mass may be detected during bimanual pelvic examination or radiologic imaging as part of a routine health maintenance examination.

For the purposes of this evidence report, we define an adnexal mass as an enlarged structure in the uterine adnexa that can either be palpated on a bimanual pelvic examination or visualized using radiographic imaging.

There are a number of conditions that can be associated with an adnexal mass. These include malignancies arising from the ovary and fallopian tube, or metastatic disease from another site (such as the breast or gastrointestinal tract), as well as a wide range of benign conditions. For the purposes of this evidence report, "management" of the adnexal mass refers to the process by which a mass is ultimately classified as benign or malignant.

The clinical significance of discriminating benign from malignant masses differs depending on the clinical setting in which the mass is initially detected. For women with symptoms, in whom surgical management may be appropriate whether or not the mass is malignant, the main reason to discriminate between benign and malignant lesions is to facilitate referral and management by clinicians who have specialized training and experience in managing ovarian malignancy, with improved outcomes. For asymptomatic women, discriminating benign from malignant disease is important both to ensure appropriate management in the setting of malignancy, but also to avoid unnecessary diagnostic procedures, including surgery, in women with asymptomatic, nonmalignant conditions.

The prevalence of malignancy may differ between women with symptomatic and asymptomatic masses, which may in turn affect the positive and negative predictive value of a test, and, potentially, sensitivity and specificity as well. Prevalence also varies with age and with family history.

This report focuses on the evidence relevant to establishing the most appropriate way to distinguish benign from malignant adnexal masses in both symptomatic and asymptomatic women. A key consideration throughout the report will be the underlying likelihood of

malignancy in the populations studied, and the impact of this prevalence on the interpretation of the results of the reviewed studies. The results of this report are intended primarily to (a) provide a resource for clinicians and policymakers developing guidelines on management of adnexal masses, and (b) provide a resource for researchers and funding agencies in identifying gaps in our knowledge and research priorities.

Methods

Working with the Agency for Healthcare Research and Quality (AHRQ), the Centers for Disease Control and Prevention (CDC), and members of the technical expert panel, we developed seven questions to be addressed, using an analytic framework which incorporated prior probability of disease, test results, and outcomes of diagnostic surgery.

We searched MEDLINE® (1966-September 2004) and the Cochrane Database of Systematic Reviews. Searches of these databases were supplemented by reviews of reference lists contained in all included articles and in relevant review articles and meta-analyses. The searches yielded a total of 1,023 citations. Pairs of readers reviewed each abstract and selected 445 articles for full text review. Specific inclusion criteria were developed for each question, and both readers were required to agree on inclusion.

We developed tables to abstract each article, and quality criteria for each question. For studies of diagnostic tests, 2-by-2 tables were constructed for each included article, and sensitivity, specificity, and positive and negative predictive values, with 95 percent confidence intervals (CIs) for each, were calculated. If not provided, we also calculated 95% CIs for articles about prevalence and adverse event rates during diagnostic surgery. For diagnostic tests, pooled estimates of sensitivity and specificity were calculated using a random-effects model.

We performed three supplemental analyses. First, we used the Nationwide Inpatient Sample (NIS), a nationally representative database containing discharge data from approximately 20 percent of U.S. hospitals. Using *International Classification of Diseases, Ninth Revision* (ICD-9) codes and the provided corrections for sample weighting, we estimated the number of cases of women 15 and older undergoing diagnostic laparoscopy and exploratory laparotomy in 2000 and 2001 for diagnoses consistent with an adnexal mass. Mortality and morbidity rates for each type of procedure within each diagnosis were also estimated.

Second, we performed a simple decision model based on serial or parallel testing using the pooled sensitivity and specificity of various tests to predict outcomes.

Finally, we used a previously developed Markov model of the natural history of ovarian cancer to explore the implications of alternative possible pathways in the development of advanced disease – specifically, that some cancers limited to the ovaries (Stage I) may spread to the upper abdomen (Stage III) without first spreading to other pelvic organs (Stage II).

Results

Question 1: *What is the prevalence of various tumor types among women with an adnexal mass, stratified by cancer status (malignant vs. benign), age, menopausal status, and size of tumor?*

In a large screening study in Kentucky the prevalence of malignant masses was 0.09 to 0.18 percent, and of benign masses 0.8 to 1.8 percent. In 16 case series, the prevalence of malignancy

ranged from 0 to 57 percent, reflecting differences in the referral patterns of the centers where the surgery was performed. The prevalences of specific types of masses also varied widely within studies. Six studies did not report the proportion of women who were postmenopausal, and none of them described whether patients were symptomatic or asymptomatic, or the type of evaluation they had undergone prior to surgery.

Question 2: *What are the sensitivity, specificity, and reliability of the bimanual pelvic examination?*

Pooled sensitivity in five studies for detection of an adnexal mass was 0.45, and pooled specificity 0.90. For distinguishing a benign from a malignant mass, pooled sensitivity in 10 studies was 0.72, specificity 0.92. When only screening studies were included, pooled sensitivity was 0.58, specificity 0.98.

Question 3: *Among women with a palpable adnexal mass on exam or a mass identified by ultrasound/imaging, what is the sensitivity/specificity of various evaluation modalities including ultrasound (transvaginal ultrasound, transabdominal ultrasound, color Doppler, two-dimensional [2D] vs. three-dimensional [3D] ultrasound), computer tomography (CT) scan, magnetic resonance imaging (MRI) scan, and CA-125 levels for distinguishing benign from malignant masses?*

A total of 153 articles were included. For morphologic appearance on ultrasound, pooled sensitivities for specific scoring systems ranged from 0.82 to 0.91, and specificities from 0.68 to 0.81. For Doppler ovarian blood flow studies, pooled (a) sensitivity and (b) specificity were: resistive index (a) 0.72 , (b) 0.90; pulsatility index (a) 0.80, (b) 0.73; maximum systolic velocity (a) 0.74, (b) 0.81; presence of vessels (a) 0.88, (b) 0.78. The combination of morphology and Doppler had pooled sensitivity of 0.86 and specificity of 0.91.

Pooled (a) sensitivities and (b) specificities of other imaging modalities were: MRI (a) 0.91, (b) 0.88; CT (a) 0.90, (b) 0.75; and positron emission tomography using an 18-Fluorodeoxyglucose tracer (FDG-PET) (a) 0.67, (b) 0.79.

Pooled sensitivity and specificity for CA-125 at a threshold of 35 U/mL were 0.78 and 0.78, respectively. In studies that compared performance by menopausal status, both sensitivity and specificity were substantially better in postmenopausal women.

Characterization of the patient population with respect to presence or absence of symptoms, or previously performed tests, was uniformly poor among studies.

Question 4: *What is the accuracy of explicit scoring systems which incorporate various combinations of imaging findings, patient risk factors, and/or CA-125 levels for detecting malignancy? Have these scoring systems been applied to a population of women before laparoscopy or laparotomy?*

We identified 36 studies. Existing validated scoring systems were all developed in mixed pre- and postmenopausal populations. The highest demonstrated specificity obtained with these scoring systems appears to be in the range of 90 to 95 percent, and, at this range of specificity, the sensitivity appears to be in the range of 65 to 80 percent. Performance was consistently worse in validation studies (done to confirm the performance of the scoring system) than in development studies. Many of the studies were applied to patients immediately prior to surgery, but the clinical presentation and prior testing were not described.

Question 5: *Among women with suspected benign masses on initial investigation, what are the sensitivity and specificity of monitoring with periodic CA-125 and/or interval ultrasound examinations for detecting malignant masses? How does the interval of testing/definition of change affect sensitivity and predictive value?*

Nine studies were identified, and, because of variable definitions and methods, no definitive conclusions could be drawn. In one large study of over 15,000 postmenopausal women, no cancers were ultimately diagnosed in a unilocular cyst less than 10 cm (2,763 women) over a mean followup of 6.3 years, although three cancers developed after resolution of the cyst or in the contralateral ovary.

Question 6: *Among women with adnexal masses, what are the morbidity and mortality from diagnostic surgery (laparoscopy or laparotomy)? At what point does the risk of surgery outweigh the risk of detecting malignancy?*

In 15 series totaling 4,915 patients, there were three deaths. Morbidity rates were also low. Comparative studies suggest lower morbidity with laparoscopy, but there is potential confounding, even in randomized studies. None of the included studies provided sufficient clinical detail to determine whether risks differed based on ultimate diagnosis.

In the NIS, both morbidity and mortality were highest in cases with a cancer diagnosis, but available codes prevented direct comparisons. In addition, because outpatient laparoscopic procedures were not included, both numerators and denominators are likely to be underestimated.

Question 7: *What are the estimated trade-offs resulting from various strategies for evaluation of the adnexal mass?*

Given the summary findings, we were unable to construct comprehensive models to estimate the likely outcomes of different strategies. In a preliminary model, serial testing with the best imaging study (morphology plus Doppler), followed by CA-125, resulted in fewer missed cancers and fewer surgeries than either test alone in postmenopausal women. Parallel testing incorporated into a scoring system resulted in slightly fewer missed cancers, but more surgeries and twice as many tests.

Because comprehensive models should ultimately include the natural history of ovarian cancer and the possible effects of screening, we identified three articles that simulated this natural history. All three assumed that ovarian cancer necessarily progresses through all four stages. Using a similarly structured model, we were able to generate estimated incidence and stage distribution similar to reported data by allowing some Stage I cancers to progress directly to Stage III. By reducing the available detection time for Stage I cancers, this would adversely affect the potential effectiveness of screening.

Discussion

Limitations of the Literature

The main limitation in the literature was the failure to adequately describe relevant patient characteristics, including the presence or absence of symptoms, and variable reporting of menopausal status. Inadequate sample size, lack of blinding, and failure to account for observer variability were also common limitations.

Limitations of the Report

The report did not include non-English publications. We did not include non-U.S. studies in our review of the prevalence of different types of adnexal mass. Given the heterogeneity of

studies, pooled estimation of sensitivity and specificity may not be appropriate. The NIS does not include outpatient procedures, and our coding algorithm may have missed some complications.

Future Research

Research priorities include: a minimal consensus data set on key patient characteristics (with results presented stratified by those characteristics); better estimates of prevalence and surgical outcomes using data sources that capture inpatient and outpatient encounters, such as Medicare or health maintenance organizations; better characterization of patient characteristics in all studies; better evidence on the value of the pelvic exam as part of routine health maintenance; and development of additional models for simulating the natural history of ovarian cancer and evaluating screening, diagnosis, and treatment strategies.

Conclusions

Developing an effective and efficient algorithm for the evaluation of any condition requires good evidence on the prevalence of the condition at the first diagnostic encounter, and the sensitivity and specificity of the potential diagnostic tests to be used. Unfortunately, the overwhelming majority of the literature we reviewed did not provide sufficient detail on important patient characteristics to allow estimation of the outcomes of different diagnostic strategies, either in the context of detecting adnexal masses or distinguishing benign from malignant masses.

All of the diagnostic tests and scoring systems we evaluated exhibited a trade-off between sensitivity and specificity – studies of a given test that reported higher sensitivity had lower specificity, and vice versa. The bimanual pelvic examination has low sensitivity for both detection of adnexal masses and discriminating benign from malignant masses, raising doubts about its utility as a screening test in asymptomatic women. In pooled analysis, the combination of ultrasound morphology and Doppler blood flow had the best combination of sensitivity and specificity, with MRI comparable. In a preliminary model, serial testing with imaging followed by CA-125 was both more sensitive and more specific than either test alone; parallel testing using both tests incorporated into the Risk of Malignancy Index resulted in fewer missed cancers (greater sensitivity) but more surgeries (lower specificity), with twice as many tests.

Studies of surgical management suffered from the same limitations in terms of description of patient characteristics, making estimation of the risks of false positive diagnostic testing impossible.

Ultimately, evaluation of potential strategies for reducing morbidity and mortality from ovarian cancer may require use of simulation models, a technique that has proven helpful in evaluating prevention strategies for other cancers. Because the natural history of ovarian cancer is relatively unknown, testing of alternative models is critical. Although a few sophisticated models exist, development of additional models would be helpful, especially in the context of evaluating results from ongoing trials of screening. If any of these trials show a benefit from screening, then the need for better evidence on the diagnostic evaluation of adnexal masses will become even more critical.

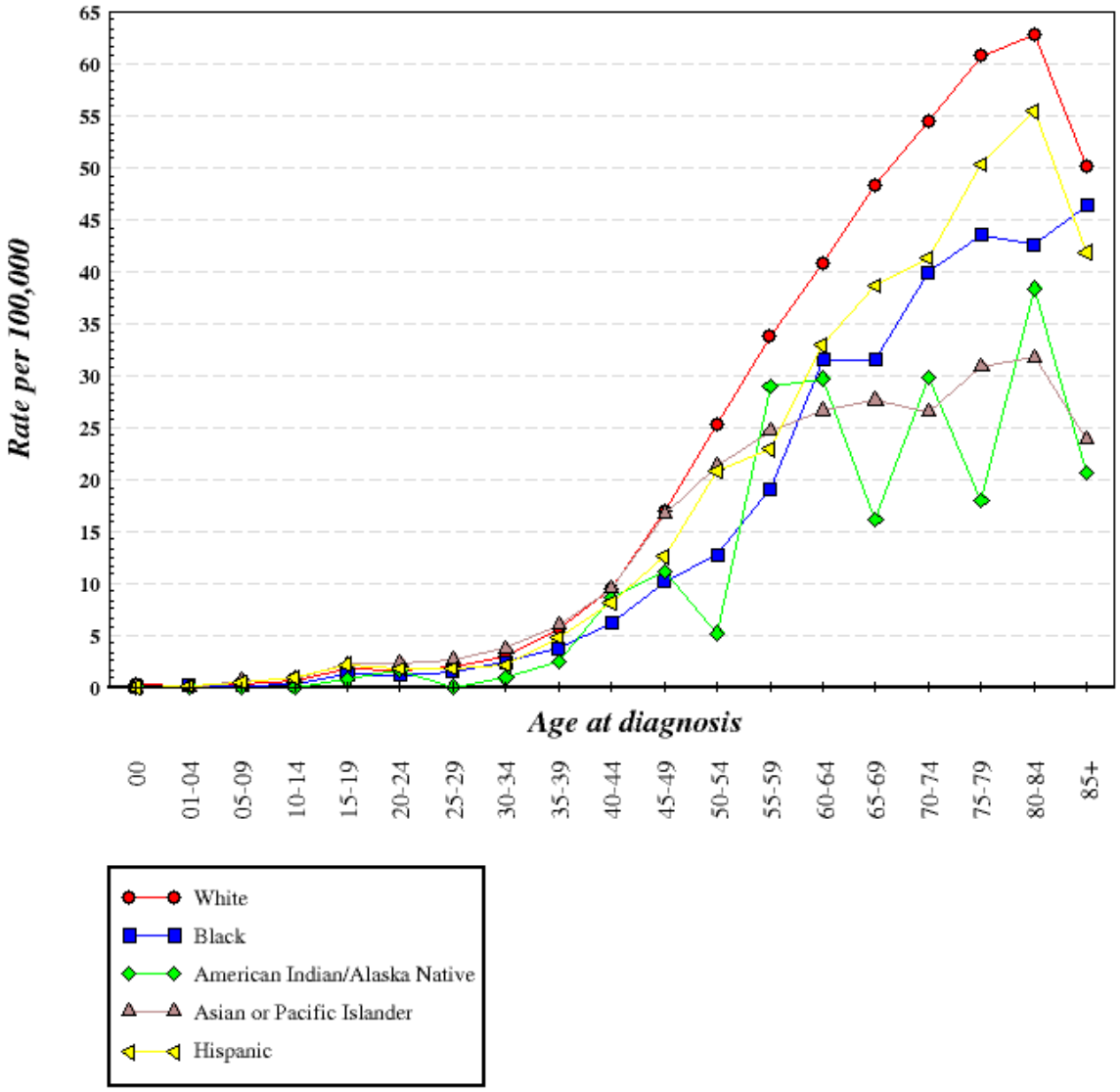
EVIDENCE REPORT

Chapter 1. Introduction

Ovarian Tumors

Cancer of the ovaries is the leading cause of cancer death from gynecologic malignancies in the United States, with an annual incidence of over 25,000 and an annual mortality of approximately 14,000.[1] Cancer incidence increases dramatically with age, being relatively rare prior to age 50 (Figure 1).

Figure 1. U.S. ovarian cancer incidence by age and race, 1992-2002



Source: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov).[2]

Ovarian cancer incidence varies by race and ethnicity. Both incidence and mortality are highest for white women (Table 1).

Table 1. Age-adjusted annual U.S. incidence and mortality per 100,000 women by race and ethnicity, 1992-2002[†]

	White	African-American	Asian/Pacific Islander	Native American	Hispanic
Incidence	15.1	10.3	10.4	8.9	11.9
Mortality	9.3	7.6	4.8	5.1	6.2

[†] Source: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov).[2]

Malignant tumors of the ovary can either arise in the ovary (primary ovarian cancer) or be the result of metastasis from another site, such as the breast or colon. Primary ovarian tumors, whether benign or malignant, can arise from three broad types of cells: the cells on the surface (epithelial cells); the cells that form eggs (germ cells); and the cells surrounding the eggs, including the cells that produce ovarian hormones (sex cord-stromal cells). Epithelial tumors are the most common type, accounting for 60 percent of all ovarian tumors and up to 90 percent of primary cancers. Sex-cord-stromal tumors account for 10 to 15 percent of all tumors, while germ cell tumors account for 25 percent of tumors. In general, sex cord-stromal tumors and germ cell tumors are relatively more common in younger premenopausal women. Thus, although ovarian cancer is relatively rare in younger women, when it does occur it is more likely to be a non-epithelial cancer than cancers in postmenopausal women.[3]

Within the broad classification of epithelial, sex cord-stromal, and germ cell tumors, tumors are further classified by the individual cell types from which they are derived. For example, the most common epithelial tumors are serous and mucinous tumors, the most common sex-cord stromal tumors are fibromas (arising from the connective tissue surrounding eggs), and the most common germ cell tumors are teratomas. Within each histological class, tumors can be benign or malignant, based on their ability to metastasize.[3]

Some epithelial tumors are classified as "borderline" or "low malignant potential" (LMP) tumors. These are tumors in which there is no invasion into the ovarian stroma, but for which histologic evidence of proliferation exists (increased cell division, changes in the appearance of the cell nucleus). There is controversy over whether these tumors represent pre-invasive cancer, and, if untreated, would go on to become a cancer, or whether they represent a subtype of tumor that has a relatively small chance of becoming a cancer.[3] In estimating the diagnostic accuracy of tests for determining whether a mass is benign or malignant, whether LMP tumors are classified as benign or malignant can have an effect on the estimates of test performance, as we will discuss later in the report.

Ovarian cancer spreads primarily by dissemination throughout the peritoneal cavity; common sites of metastasis are the small and large bowel, the omentum, the liver, and the diaphragm. Spread to retroperitoneal lymph nodes is also common.

Treatment for ovarian cancer consists of surgical removal of the ovaries, fallopian tubes, and uterus (if present), along with as much metastatic disease as possible; if there is no obvious spread beyond the ovaries, the lymph nodes are sampled to determine if there has been lymphatic metastasis. Surgery is followed by chemotherapy, with responsiveness to chemotherapy depending on the amount of tumor left after surgical removal and the cell type of tumor, among other factors.[3]

The high case-fatality rate observed in ovarian cancer has largely been attributed to the fact that most ovarian cancers are diagnosed in advanced stages (Stage III, where the cancer has spread beyond the pelvis to organs of the upper abdominal cavity, and Stage IV, where the cancer has spread outside of the peritoneal cavity), when survival is poor. Stage I cancer (limited to the ovaries) has a survival rate of over 90 percent. Thus, there has long been an emphasis on early detection of ovarian cancer, in the belief that detection in early stages will lead to decreases in morbidity and mortality, just as cervical cancer screening has resulted in substantial reductions in morbidity and mortality from cervical cancer. The detection of a mass in the area of the ovaries and fallopian tubes (the uterine adnexae) raises the possibility of ovarian cancer, which necessitates further study to rule out malignancy.

This evidence report was prepared by the Duke Evidence-based Practice Center, in partnership with the Centers for Disease Control and Prevention (CDC) and the Agency for Healthcare Research and Quality (AHRQ). The purpose of the report is to provide followup data regarding key issues identified at two conferences sponsored by CDC, one in November 2000 on broad issues in preventing morbidity and mortality from ovarian cancer,[4] and one in May 2002 on the use of ultrasound in the diagnosis of ovarian cancer.[5]

Definition of an Adnexal Mass

For the purposes of this report, we define an adnexal mass as an enlarged structure in the uterine adnexa which can either be palpated on a bimanual pelvic examination or visualized using radiographic imaging. The normal ovary is approximately 3 cm in length, decreasing in size after menopause.[6] In terms of physical examination, the precise size definition used in the literature is quite variable and, in practice, may also vary depending on the ease with which the examination is performed, the patient's body habitus, the examiner's experience, the time taken during the exam, and the presence of other abnormalities such as uterine fibroids. Historically, because of the decrease in size after menopause, any palpable mass in a postmenopausal woman has been considered abnormal (the "palpable postmenopausal ovary syndrome").[7] As discussed below, some masses may ultimately prove to not be ovarian in origin.

The definition of an abnormal structure on radiologic imaging is also quite variable. Small fluid-filled cysts are quite common in both pre- and postmenopausal women. For the purposes of this report, we consider any structure observed during radiologic imaging that prompts additional evaluation (such as measurement of serologic markers or further imaging) as a mass.

Detection of an Adnexal Mass

There are three main clinical routes by which an adnexal mass may be detected. First, women with symptoms may have an adnexal mass detected as part of their evaluation for those symptoms, either by physical exam or radiographic imaging. Because ovarian cancer often presents with vague abdominal symptoms, we would consider any evaluation for symptoms to be in symptomatic women. Second, the mass may be detected as part of a routine health maintenance examination. Finally, it is possible that an asymptomatic mass could be detected during imaging done for another indication. In premenopausal women, the most likely scenario where this would occur would be during ultrasound evaluation during pregnancy. Another common scenario in peri- or postmenopausal women would be evaluation for uterine bleeding;

because uterine bleeding is not a common symptom of ovarian cancer, a finding of an adnexal mass during evaluation for bleeding could be considered as an incidental finding. Because malignancy is rare during pregnancy, and because the technical considerations for both diagnosis and management are different, the most appropriate management of masses detected during pregnancy, especially if detected serendipitously by ultrasound, is outside of the scope of this report.

We did not identify any literature that would allow an estimate of the proportions of women with adnexal masses presenting by each route; as we will discuss, this is a major deficiency of the literature. The proportions are likely to vary by setting, referral patterns, patient thresholds for seeking care, physician thresholds for diagnostic tests, and other factors. For example, one gynecologic oncologist estimated that well over half of the referrals for evaluation in a large health maintenance organization were for incidentally detected masses (W. Kinney, personal communication).

Types of Adnexal Mass

Conditions that can present as an adnexal mass include:

- Benign primary ovarian tumors – epithelial, sex cord-stromal, and germ cell;
- Borderline and malignant ovarian tumors – epithelial, sex cord-stromal, and germ cell;
- Metastatic malignant tumors – most commonly breast and gastrointestinal tract;
- Masses arising from the fallopian tube – most commonly benign, including hydrosalpinx (a large, fluid-filled fallopian tube) and pyosalpinx (an infected, pus-filled fallopian tube); primary fallopian tube malignancies can occur, but are relatively rare.
- Masses arising from the uterus – most commonly benign leiomyomas (fibroids);
- Masses arising from the gastrointestinal tract – diverticula of the colon, large colonic tumors, tumors of the appendix;
- Masses arising from the urinary tract – pelvic kidneys, diverticula of the ureter;
- Masses arising from remnants of embryological development;
- Endometriosis;
- Pelvic inflammatory disease;
- Cysts arising from normal ovarian functions, such as development of eggs (follicular cysts) and ovulation (corpus luteum cysts).

Management of the Adnexal Mass

With such a wide range of potential causes, and with a wide range of appropriate therapeutic options, precise diagnosis of a mass, especially in symptomatic women, is important. Once diagnosed, a mass may be managed in a variety of ways, ranging from observation to surgical removal and chemotherapy. However, a review of the test characteristics of various methods for obtaining precise diagnoses of specific conditions, and of the range of medical and surgical treatment options for each condition, is beyond the scope of this report. For our purposes, “management” of the adnexal mass refers to the process by which a mass is ultimately classified as benign or malignant.

Importance of Discriminating Benign from Malignant Masses

The clinical significance of discriminating benign from malignant masses differs depending on the clinical setting in which the mass is initially detected.

In women who initially present with symptoms, diagnosis of the underlying cause of the mass is important since it may help define available treatment options. Although medical therapy may relieve symptoms in some cases, surgical management is the treatment of choice for many conditions. Because surgery may ultimately be the most appropriate management for symptomatic adnexal masses, the main reason to discriminate between benign and malignant lesions is to facilitate referral and management by clinicians with specialized training and experience in managing ovarian malignancy, with improved outcomes.[8-10]

The other main group of women with adnexal masses consists of those without symptoms who have a mass detected through either physical examination or imaging. No organization currently recommends routine screening with serum markers or imaging for ovarian cancer.[11,12] The U.S. Preventive Services Task Force gives screening (including serum markers, imaging, or pelvic examination) a "D" recommendation (fair evidence against screening).[13] However, because an annual pelvic examination continues to be recommended by professional organizations such as the American College of Obstetricians and Gynecologists (ACOG),[11,14] many asymptomatic women may have an adnexal mass detected during a periodic health maintenance examination. In this setting, discriminating benign from malignant disease is important not only to ensure appropriate management in the setting of malignancy, but also to avoid unnecessary diagnostic procedures, including surgery, and anxiety in women with asymptomatic, nonmalignant conditions. In some cases, there may be a rationale for removing certain asymptomatic benign lesions, including prevention of malignant transformation; prevention of ovarian torsion (a condition where the ovary twists and occludes its blood supply, causing abdominal pain and possibly resulting in loss of ovarian function); prevention of rupture, which might lead to acute symptoms or a worse prognosis (for example, in the case of endometriosis); prevention of more advanced or complicated surgery for a larger mass or more extensive pathologic process after the development of symptoms; and, for premenopausal women, possible enhancement of fertility. A review of the evidence (or lack of evidence) supporting these rationales is beyond the scope of this report.

Significance of Clinical Presentation in Evaluation of Management Strategies

As discussed above, the results of tests used to distinguish benign from malignant disease have different implications depending on whether the patient is symptomatic or asymptomatic. However, clinical presentation also has implications for interpretation of test results.

Diagnostic or screening tests are most commonly characterized by their sensitivity and specificity. The sensitivity of a test is the probability that, given the underlying presence of the disease, the test result will be positive; 100 percent minus the sensitivity is commonly called the false negative rate. The specificity of the test is the probability that, given the underlying absence of disease, the test result will be negative; 100 percent minus the specificity is

commonly called the false positive rate. In an ideal evaluation, the sensitivity and specificity of the test are independent of the underlying probability, or prevalence, of disease.

Clinically, the more common scenario is that the clinician is aware of the test result and needs to know the probability of the presence or absence of disease. In this setting, the positive and negative predictive values of the test are more important.

The negative predictive value of a test is the probability that, given a negative test result, the patient truly does not have disease. It is a function of three parameters: the pretest probability of the disease, the sensitivity of the test, and the specificity of test:

$$\frac{(1 - \textit{Prevalence}) * \textit{Specificity}}{[(1 - \textit{Prevalence}) * \textit{Specificity}] + [\textit{Prevalence} * (1 - \textit{Sensitivity})]}$$

As can be seen in the equation, the negative predictive value is much more dependent on test sensitivity than test specificity. Negative predictive value will be high when test sensitivity is high, and when prevalence is low (i.e., disease is rare).

Similarly, the positive predictive value is the probability that, given a positive test result, the patient actually has the disease. It is also a function of prevalence, sensitivity, and specificity:

$$\frac{\textit{Prevalence} * \textit{Sensitivity}}{(\textit{Prevalence} * \textit{Sensitivity}) + [(1 - \textit{Prevalence}) * (1 - \textit{Specificity})]}$$

Positive predictive value is high when a test has high specificity, or when prevalence is high (disease is common).

For any given test, the positive predictive value will be higher and the negative predictive value lower when used in populations where the disease is common compared to populations where the disease is rare, while the positive predictive value will decrease and the negative predictive value increase as the disease becomes less common. This effect of prevalence on predictive values is independent of test sensitivity and specificity. The significance of the prevalence of disease in the population in which test characteristics are being evaluated is even more critical because, under some types of study design, disease prevalence can also affect estimates of sensitivity and specificity.[15]

Therefore, variations in the prevalence of malignancy among women with different clinical presentations will affect at least predictive values, and possibly sensitivity and specificity estimates. The prevalence of ovarian cancer clearly rises with age, so age and/or menopausal status are important considerations in evaluating management strategies in both the symptomatic and asymptomatic patient with an adnexal mass.

The prevalence of malignancy among asymptomatic women with an adnexal mass will be a function of the underlying prevalence or incidence of malignancy and the test characteristics of the initial test used to detect the mass. Evaluation of the different screening tests and strategies for early detection of ovarian cancer is beyond the scope of this report, especially since there are at least three large trials still ongoing.[16-18] However, in order to properly interpret the results of tests performed in asymptomatic women with pelvic masses, some estimate of the underlying probability of malignancy among these women is needed. Since many of these women are likely identified through a bimanual pelvic examination, deriving this estimate requires an assessment of the sensitivity and specificity of the pelvic examination. Symptomatic patients may be more likely to have an underlying adnexal malignancy, especially among postmenopausal women.[19]

In any series of women with adnexal masses, the proportion of women who are symptomatic and asymptomatic will likely determine the prevalence, and thus the predictive values of the diagnostic tests used to evaluate the mass.

Summary

In summary, this report focuses on the evidence relevant to establishing the most appropriate way to distinguish benign from malignant adnexal masses in both symptomatic and asymptomatic women. A key consideration throughout the report will be the underlying likelihood of malignancy in the populations studied, and the impact of this prevalence on the interpretation of the results of the reviewed studies. The results of this report are intended primarily to (a) provide a resource for clinicians and policymakers developing guidelines on management of adnexal masses, and (b) provide a resource for researchers and funding agencies in identifying gaps in our knowledge and research priorities.

Chapter 2. Methods

This section of the report describes the basic methodology used to develop the evidence report, including topic assessment and refinement, analytic framework, literature search strategies and results, literature screening and grading process and criteria, data abstraction and analysis methods, and quality control procedures.

Topic Assessment and Refinement

The Centers for Disease Control and Prevention (CDC) and the Agency for Healthcare Research and Quality (AHRQ) originally identified five key questions to be addressed by the report, focused on management of adnexal masses in peri- and postmenopausal women. The Duke research team clarified and refined the overall research objectives and key questions by first consulting with the two study sponsors, AHRQ and CDC, at which time two questions were added, and then by convening a panel of national experts who would serve as advisors to the project. These experts were selected to represent relevant specialties including radiology, obstetrics-gynecology, and gynecologic oncology, as well as national professional societies, including the American College of Obstetricians and Gynecologists (ACOG), the Society of Gynecologic Oncologists (SGO) and the American College of Radiology (ACR). Members of the technical expert panel were:

Susan Ascher, MD; Department of Radiology, Georgetown University Hospital; Washington, DC (ACR)

Michael L. Berman, MD; Division of Gynecologic Oncology, UCI Medical Center; Orange, CA (SGO)

Barry B. Goldberg, MD; Department of Radiology, Thomas Jefferson University Hospital; Philadelphia., PA (ACR)

Edward E. Partridge, MD; Department of Obstetrics and Gynecology, University of Alabama, Birmingham; Birmingham, AL (American Cancer Society)

George F. Sawaya, MD; Department of Obstetrics and Gynecology, University of California, San Francisco; San Francisco, CA

Howard T. Sharp, MD; University of Utah Hospitals and Clinics; Salt Lake City, UT (ACOG)

Stanley Zinberg, MD, MS; ACOG; Washington, DC

As a result of an initial conference call with the technical experts, AHRQ, and CDC, the Duke research team modified the key research questions originally proposed in the Task Order in two fundamental ways: (1) The questions were expanded to include women of all ages, and (2) Question 6 would include laparotomy data, where available. After review of a draft version of

the report by the technical experts and additional reviewers, the order of the questions was also changed to allow a more logical flow.

The key questions addressed by this report are:

Question 1: What is the prevalence of various tumor types among women with an adnexal mass, stratified by cancer status (malignant vs. benign), age, menopausal status, and size of tumor?

Question 2: What are the sensitivity, specificity, and reproducibility of the bimanual pelvic examination?

Question 3: Among women with a palpable adnexal mass on exam or a mass identified by ultrasound/imaging, what is the sensitivity/specificity of various evaluation modalities including ultrasound (transvaginal ultrasound, transabdominal ultrasound, color Doppler, two-dimensional [2D] vs. three-dimensional [3D] ultrasound), computer tomography (CT) scan, magnetic resonance imaging (MRI) scan, and cancer antigen 125 (CA-125) levels for diagnosing malignant masses?

Question 4: What is the accuracy of explicit scoring systems which incorporate various combinations of imaging findings, patient risk factors, and/or CA-125 levels for detecting malignancy? Have these scoring systems been applied to a population of women before laparoscopy or laparotomy?

Question 5: Among women with suspected benign masses on initial investigation, what are the sensitivity and specificity of monitoring with periodic CA-125 and/or interval ultrasound examinations for detecting malignant masses? How does the interval of testing/definition of change affect sensitivity and predictive value?

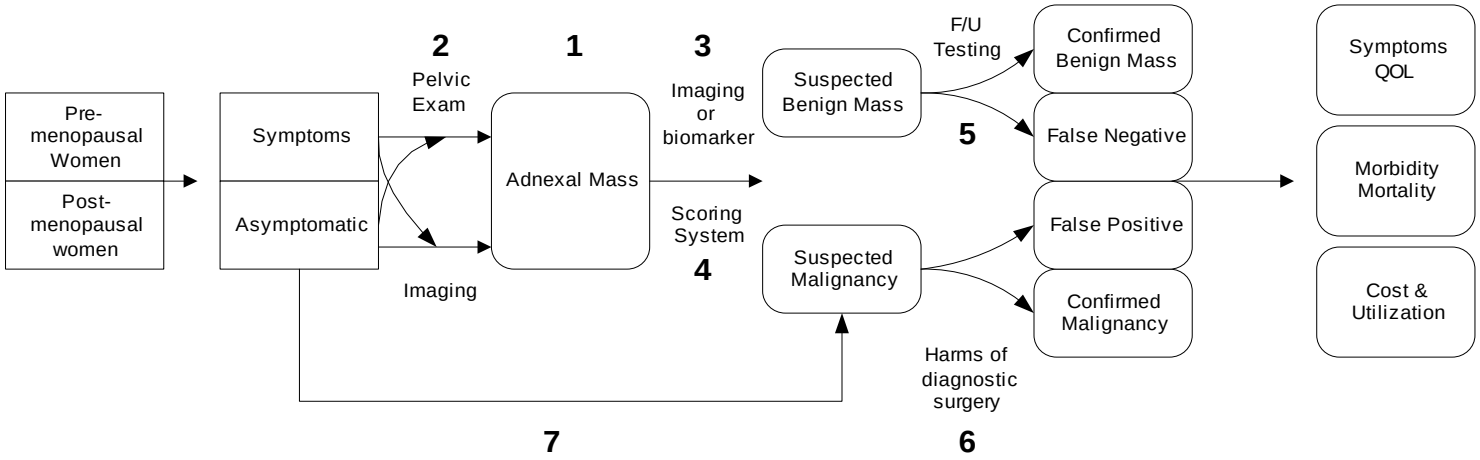
Question 6: Among women with adnexal masses, what are the morbidity and mortality from diagnostic surgery (laparoscopy or laparotomy)? At what point does the risk of surgery outweigh the risk of detecting malignancy?

Question 7: What are the estimated trade-offs resulting from various strategies for evaluation of the adnexal mass?

Analytic Framework

Based on the original proposal and discussions with CDC, AHRQ, and the technical expert panel, we developed the following analytic framework to structure our review and synthesis (Figure 2).

Figure 2. Analytic framework for evidence report (numbers refer to key questions)



Comments on this analytic framework are as follows:

- Separate consideration of age or menopausal status is important, since several factors that may affect the probability that a given adnexal mass is malignant may vary with age and/or menopausal status: the underlying incidence of various conditions that result in an adnexal mass, the frequency of contact with clinicians, the type and length of followup, and the prevalence of other conditions that may cause symptoms similar to those caused by ovarian malignancy or other symptomatic pelvic pathology. Race/ethnicity may also play a role, both in the relative likelihood of malignancy and the likelihood of other conditions and contact with clinicians.
- A variety of conditions, both benign and malignant, can cause a mass in the adnexa. The underlying prevalence of each type of condition, along with the sensitivity and specificity of the initial diagnostic test, will determine the proportion of patients with a given test result who are truly disease-free, or who truly have disease. The evidence on the prevalence of these conditions is reviewed in Question 1.
- Women can present with an adnexal mass in one of two ways – through presentation with symptoms and subsequent detection of a mass through a physical examination, or through detection of a mass in an asymptomatic woman during physical examination or an imaging study. The ultimate probability of malignancy may vary based on how an adnexal mass is initially detected, since the prevalence of malignancy at this stage will drive the positive and negative predictive values of all subsequent tests. Because many women will initially have their masses detected through a bimanual pelvic examination, we review the evidence on the sensitivity and specificity of this component of the physical examination in Question 2.
- After the initial diagnosis of an adnexal mass, the choice of the next test will provide a revised estimate of the probability of a given disease. Although determining this probability is important in the symptomatic patient so that she may receive appropriate therapy, it is even more important in the asymptomatic patient, who runs the risk of undergoing unnecessary surgery for a benign condition if the test is falsely positive. Question 3 addresses the sensitivity and specificity of tests commonly used as "next step" diagnostic procedures.
- Frequently, a combination of various test results and patient characteristics can provide better discrimination between diseased and non-diseased, or benign and malignant, than any single test parameter. Question 4 addresses the performance of various multivariate scoring systems in discriminating benign from malignant masses.
- Because 100 percent sensitivity is difficult to achieve, some tests will be falsely negative. One strategy to minimize the consequences of a false negative test would be to monitor the patient with a specified test or tests, at a specified frequency, for a specified duration. Question 5 addresses the evidence for the effectiveness of such an approach, and which combination of test, test frequency, and duration of followup offers optimal performance.
- The ultimate diagnosis of ovarian malignancy requires surgical exploration, either through laparoscopy or laparotomy. Although an adverse outcome of surgery is not desirable under any circumstances, patients who undergo surgery because of a symptomatic mass have the possibility of improvement in symptoms, while, for patients who ultimately prove to have an ovarian malignancy, surgical management with adequate staging and reduction in tumor bulk appears to improve outcomes. However, for patients

with some asymptomatic benign masses, the benefits of surgery may be less clear while providing substantial risks. Question 6 addresses the risks of diagnostic surgery, both laparoscopy and laparotomy, for women with adnexal masses.

- Finally, estimating the benefits, harms, and costs of various management strategies, including screening, for ovarian cancer is complex. Synthesizing the wide range of data and incorporating uncertainty, as well as missing data, can often be done using simulation models. Question 7 presents an initial attempt at summarizing the likely outcomes of several different diagnostic strategies. Because modeling the natural history of ovarian cancer will ultimately be important for comprehensive analyses of different screening and diagnostic strategies, we also review existing models for the natural history of ovarian cancer with special attention paid to underlying assumptions.

Literature Search and Review

Sources

The primary sources of literature were MEDLINE® (1966-September 2004) and the Cochrane Database of Systematic Reviews. Searches of these databases were supplemented by reviews of reference lists contained in all included articles and in relevant review articles and meta-analyses.

Search Strategies

The basic search strategy used the National Library of Medicine's Medical Subject Headings (MeSH) key word nomenclature developed for MEDLINE® and was adapted for use in the other databases. The searches were limited to the English language. The texts of the three major search strategies are given in Appendix A.* The searches yielded a total of 677 citations, whose records are maintained in a ProCite[20] database.

Abstract and Full-text Screening

Paired researchers from the Duke research team independently reviewed a set of abstracts and classified each as "include" or "exclude" according to study-specific criteria, which they developed. An abstract was included if at least one of the paired reviewers recommended that it be included. A total of 445 abstracts were included for the further "full-text review" stage. Interrater reliability for include/exclude decisions was tested by having 10 pairs of readers review 138 abstracts. Agreement was good to excellent (kappa 0.66 to 0.95).

At the full-text review stage, the paired researchers independently reviewed a set of the articles, and indicated a decision to "include" or "exclude" the article for the data abstraction stage. When a pair of reviewers arrived at a different opinion about whether to include an article, they were asked to reconcile the difference. Detailed inclusion and exclusion screening criteria were developed by research question and are listed below.

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/adnexal/adnexal.pdf.

Full-text Screening Criteria

Initially, the patient population was limited to peri- and postmenopausal women, and only articles that provided data specifically by age or menopausal status were included. After initial discussion with the expert panel, the search was expanded to include premenopausal women.

Question 1. Background clarifications were as follows:

(1) The search should be limited to (a) screening studies and (b) case series of women with an undiagnosed mass (not just women who went to laparoscopy/path diagnosis).

(2) Pathology list:

a. Benign
   i. Uterine leiomyoma
   ii. Nonneoplastic cysts, such as:
      1. Follicular (functional) cysts
      2. Corpus luteal (functional) cysts
      3. Theca lutein cysts
      4. Simple cysts
      5. Peritoneal inclusion cysts
      6. Paraovarian cysts
      7. Hemorrhagic cysts
      8. Endometrial cyst
   iii. Polycystic ovary disease
   iv. Cystic teratoma (dermoid cyst)
   v. Hydrosalpinx,
   vi. Cystadenoma
   vii. Fibroma

b. Malignant ovarian neoplasms
   i. Adenocarcinoma
   ii. Others

c. Tumors of low malignant potential

Screening criteria for Question 1 were:

(1) undiagnosed mass (regardless of whether symptomatic or asymptomatic; detected by palpation or ultrasound imaging);
(2) exclude if $n < 50$; if $n \geq 50$, write n on decision sheets;
(3) histology diagnosis;
(4) screened women without mass (case series or cohort) or women with adnexal mass (case series).

Question 2. Screening criteria were as follows:

(1) comparison of bimanual pelvic examination to a reference standard;
(2) $n \geq 20$;
(3) able to construct 2-by-2 table for test characteristics.

Question 3. Screening criteria were as follows:

(1) undiagnosed mass (regardless of whether symptomatic or asymptomatic; detected by palpation or ultrasound imaging) or screening population;
(2) disease status distinguishes malignant from non-malignant;
(3) must have 20 or more subjects;

(4) disease status must be verified by histology or negative surgery (laparoscopy/laparotomy);
(5) test is ultrasound, CT, MRI, PET, serum CA-125, or bimanual pelvic exam;
(6) able to construct 2-by-2 table for test characteristics.

Question 4. Screening criteria were as follows:
(1) patients with cancer;
(2) studies with scoring, risk score, combined modality approach;
(3) assesses predictive value of two or more variables (radiographic, patient characteristics or CA-125) using multivariable model;
(4) screening studies;
(5) $n \geq 50$.

Question 5. Screening criteria were as follows:
(1) $n \geq 50$;
(2) histology or followup interval = at least 9 months;
(3) outcome = continued negative test with no clinical evidence of developing ovarian cancer.

Question 6. Screening criteria were as follows:
(1) procedure = operative laparoscopy for adnexal mass, with or without biopsy;
(2) addresses complications of procedure (morbidity or mortality);
(3) $n \geq 100$ for morbidity.

Question 7. Screening criterion was as follows: article described mathematical or computer model of natural history of ovarian cancer.

Summaries of the results of the abstract screening and full-text review are provided in Tables 2 and 3. A list of excluded articles by reason for exclusion is found in Appendix B.*

Table 2. Results of abstract screening and full-text review

Articles identified	1,023
Abstracts reviewed	1,023
Included	445
Excluded	578
Full-text articles reviewed	445†
Included	204
Excluded	269

† The combined number of included (204) and excluded (269) articles exceeds the total 445 reviewed at the full-text level because 28 articles were considered excluded for one question, but included for another question.

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

Table 3. Included full-text articles by research question

Question	Number of articles
Question 1: Prevalence of tumor types	20
Question 2: Bimanual pelvic examination	14
Question 3: Single modality tests	153
Question 4: Explicit scoring systems	36
Question 5: Monitoring women with suspected benign masses	9
Question 6: Surgical morbidity and mortality	24
Question 7: Modeling diagnostic strategies	4
Total number of included articles	204[†]

[†] Some articles were included for more than one question.

Data Abstraction and Development of Evidence Tables

The Duke research team developed and piloted evidence table formats for abstracting data to answer each of the seven research questions (see Appendix C[*]). Based on clinical expertise, a pair of researchers was assigned to one of the seven research questions to abstract the data from the eligible articles. One of the paired researchers abstracted the data into the evidence tables, and the second researcher over-read the article and accompanying evidence table to check for accuracy and completeness. The completed evidence tables are provided in Appendix D.[*]

Quality Assessment Criteria

At the data abstraction stage, the researcher was asked to evaluate each included article for factors affecting internal and external validity. The quality assessment criteria varied by question and are listed below. Researchers were instructed to assign a + or - to each item, and provide a brief rationale for each decision.

Quality criteria were as follows:

Question 1: *What is the prevalence of various tumor types among women with an adnexal mass, stratified by cancer status (malignant vs. benign), age, menopausal status, and size of tumor?*

- Size of population from which sample drawn. Rationale: Ideally, data on prevalence would come from population-based studies; alternatively, a precise description of the population served by a given center (for case series) allows comparison to other studies. Credit given for description.
- Number of cases. Rationale: Small numbers, especially in the denominator, decrease the precision of the estimate of proportion/prevalence.
- Patient selection. Rationale: The process by which patients come to undergo surgery may affect the prevalence of underlying disease, or the proportion of different types. For example, if one group of patients was more likely to undergo medical treatment for certain types of adnexal findings (such as oral contraceptives for possible functional

[*] Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

cysts), the distribution of types of masses would be different than in a group less likely to undergo medical therapy prior to surgery. Studies were given credit if this process were described.

- Application of reference standard. Rationale: If either all or a random sample of test negative subjects do not undergo the reference standard test, significant overestimation of test sensitivity can occur. Studies given credit if all patients underwent reference standard; alternative standards (such as pre-defined followup) were allowed for screening studies.[15,21]

Question 2: *What are the sensitivity, specificity, and reliability of the bimanual pelvic examination?*

- Reference standard. Rationale: Histology or, at the least, intraoperative visualization, is the recognized reference standard for ovarian or other adnexal pathology. Studies given credit if all subjects underwent this reference standard (documented followup acceptable in screening studies).
- Verification bias. Rationale: If either all or a random sample of test negative subjects do not undergo the reference standard test, significant overestimation of test sensitivity can occur. Studies given credit if all patients underwent reference standard; alternative standards (such as predefined followup) were allowed for screening studies.
- Test reliability/variability. Rationale: Inter- and intraobserver variability can, at least, affect the precision of estimates of test characteristics (if random), or can bias results in one direction or the other (if systematic). Studies given credit if test reliability/variability were measured, other studies measuring it were referenced, or if it was discussed as an issue.
- Sample size. Rationale: Small sample sizes limit the precision of estimates, particularly for test characteristics, which are proportions. Studies given credit if sample size discussed, or if study over 100 subjects.
- Statistical tests. Rationale: Inappropriate use of statistical tests (e.g., use of parametric tests for nonparametric data) or inappropriate interpretation of results (concluding no difference for underpowered studies) can lead to invalid conclusions about a study. Studies given credit if no examples of inappropriate use identified.
- Blinding. Rationale: Awareness of other relevant information (such as clinical history or, in the case of retrospective studies where images are reviewed outside of the clinical setting, the ultimate diagnosis) can lead to biased interpretation of results. Studies given credit if blinding explicitly described.
- Definition of +/- on screening test. Rationale: The ability to replicate a study, or to compare results between studies, depends on a description of the criteria for defining a positive test. Studies given credit if definition provided, or reference for definition provided.

Question 3: *Among women with a palpable adnexal mass on exam or a mass identified by ultrasound/imaging, what is the sensitivity/specificity of various evaluation modalities including ultrasound (transvaginal ultrasound, transabdominal ultrasound, color Doppler, 2D vs. 3D ultrasound), CT scan, MRI scan, and CA-125 levels for diagnosing malignant masses?*

- Reference standard
- Verification bias
- Test reliability/variability
- Sample size

- Statistical tests
- Blinding
- Definition of +/- on screening test

Rationale for these criteria is the same as for Question 2.

Question 4: *What is the accuracy of explicit scoring systems which incorporate various combinations of imaging findings, patient risk factors, and/or CA-125 levels for detecting malignancy? Have these scoring systems been applied to a population of women before laparoscopy or laparotomy?*

- Reference standard. Rationale: Histology or, at the least, intraoperative visualization, is the recognized reference standard for ovarian or other adnexal pathology. Studies given credit if all subjects underwent this reference standard (documented followup acceptable in screening studies).
- Verification bias. Rationale: If either all or a random sample of test negative subjects do not undergo the reference standard test, significant overestimation of test sensitivity can occur. Studies given credit if all patients underwent reference standard; alternative standards (such as pre-defined followup) were allowed for screening studies.
- Test reliability/variability. Rationale: Inter- and intra-observer variability can, at least, affect the precision of estimates of test characteristics (if random), or can bias results in one direction or the other (if systematic). Studies given credit if test reliability/variability were measured, other studies measuring it were referenced, or if it was discussed as an issue.
- Sample size. Rationale: Small sample sizes limit the precision of estimates, particularly for test characteristics, which are proportions. In the setting of multivariate models, study power is limited by the number of cases in the data set. Studies given credit if sample size discussed.
- Statistical tests. Rationale: Inappropriate use of statistical tests (e.g., use of parametric tests for nonparametric data) or inappropriate interpretation of results (concluding no difference for underpowered studies) can lead to invalid conclusions about a study. Studies given credit if no examples of inappropriate use identified.
- Blinding. Rationale: Awareness of other relevant information (such as clinical history or, in the case of retrospective studies where images are reviewed outside of the clinical setting, the ultimate diagnosis) can lead to biased interpretation of results. Studies given credit if blinding explicitly described.
- Definition of +/- on screening test. Rationale: The ability to replicate a study, or to compare results between studies, depends on a description of the criteria for defining a positive test. Studies given credit if definition provided, or reference for definition provided.
- Explicit validation method. Rationale: A scoring system will often perform differently when tested in a data set other than the one in which it was developed. Studies given credit if the method for validating the system was explicitly described or referenced.

Question 5: *Among women with suspected benign masses on initial investigation, what are the sensitivity and specificity of monitoring with periodic CA-125 and/or interval ultrasound examinations for detecting malignant masses? How does the interval of testing/definition of change affect sensitivity and predictive value?*

- Reference standard
- Verification bias

- Test reliability/variability
- Sample size
- Statistical tests
- Blinding
- Definition of +/- on screening test

Rationale for these criteria is the same as for Question 2.

Question 6: *Among women with adnexal masses, what are the morbidity and mortality from diagnostic surgery (laparoscopy or laparotomy)? At what point does the risk of surgery outweigh the risk of detecting malignancy?*

- Size of population from which sample drawn
- Number of cases
- Patient selection
- Application of reference standard

Rationale for these criteria is the same as for Question 1.

Question 7: *What are the estimated trade-offs resulting from various strategies for evaluation of the adnexal mass?*

For this question, we examined published models of ovarian cancer and qualitatively assessed the underlying assumptions and evidence for them.

Additional Analyses

Test Characteristics and Confidence Intervals

For test characteristics, a Microsoft Excel® spreadsheet was developed which calculated appropriate test characteristics (sensitivity, specificity, negative predictive value, positive predictive value) for individual studies if studies provided enough data to input (a) values for individual cells of a 2-by-2 table, (b) the prevalence of disease and values for sensitivity and specificity, or (c) sufficient data to solve for two equations involving sensitivity, specificity, or predictive values. Ninety-five percent confidence intervals were automatically estimated using the approximate formula for proportions:

$$p \pm 1.96 * \sqrt{p * (1 - p) / N}$$, where p = point estimate of proportion, N = total sample size.

Prevalence and Event Rates and Confidence Intervals

For Questions 3 and 6, prevalence of different mass types, and morbidity and mortality rates, were also calculated using the above formula. For studies where the numerator of a particular proportion was 0, the upper bound was estimated using the formula:

$$p + 2.56 * \sqrt{p * (1 - p) / N}$$, where p = 2/(N + 2).

Estimation of Pooled Sensitivity and Specificity

For Questions 2, 3, and 4, we used two complementary methods for assessing diagnostic test performance: (1) summary receiver operating characteristic (ROC) analysis; and (2) independently combined sensitivity and specificity values. We calculated pooled sensitivity and specificity estimates, along with 95 percent confidence intervals and summary ROC curves, using Meta-Stat 0.6, a shareware program for performing meta-analyses of diagnostic tests.[22] In this software, logits of sensitivity and specificity values are pooled, using a random-effects model weighted by the inverse of the variance.[23]

We combined the sensitivity and specificity values of the tests across studies using a random-effects model to estimate the average values. A random-effects model incorporates both the within-study variation (sampling error) and between-study variation (true treatment-effect differences) into the overall treatment estimate. It gives a wider confidence interval than the fixed-effect model (which considers only within-study variability) when estimates are based on heterogeneous results.

When each is combined separately, sensitivity and specificity tend to underestimate the true test sensitivity and specificity; however, they can provide an indication of the approximate test operating point for most of the studies.

Summary ROC curves are a potentially useful graphical summary of the diagnostic test performance data. In brief, each study provides a pair of sensitivity and specificity values to the analysis. After logistic transformation of data, a linear model is fitted to the observed studies using regression analysis. This best-fit model can then be transformed back to ROC space and plotted as curve. A summary ROC curve can be thought of as an ROC curve that describes joint changes in sensitivity and specificity with changes in cutoff values. The ideal position of an ROC curve is near the upper left corner. The area under the curve (AUC) is another summary measure of the degree of discrimination of a test.

The summary ROC method assumes that the variability in the reported sensitivity and specificity values from different studies is due to different cutoff values (explicit or implicit) being applied.[24] However, the summary ROC curve can summarize studies whose variability may be due to other sources of variation, since the summary ROC curve no longer ties specific cutoff values to specific intervals of the curve. One can think of a summary ROC curve as an overall estimate of the discrimination ability of a test.

When there is little variability in the test results – i.e., when studies appear to be operating at similar thresholds and report similar results – summary ROC analysis provides little additional information. In this case, separately averaged sensitivity and specificity values across studies will give similarly useful summary information. However, where there is substantial variability in test results, the separately averaged sensitivity and specificity values tend to have wide confidence intervals and have means that do not characterize any of the studies. In this case, SROC curves provide a more suitable analysis framework.

Estimates of National Rates of Surgery for Adnexal Mass

The Nationwide Inpatient Sample (NIS) is a public access database maintained by AHRQ. The NIS represents a stratified sample of approximately 20 percent of all discharges from U.S. hospitals; data for the year 2000 contain administrative discharge data from hospitals in 28 states, while 2001 contains data from 33 states.[25] Weights are provided in order to allow

estimation of national data based on this sample. We used data from 2000 and 2001 to provide supplemental data on the frequency of diagnostic laparoscopy and exploratory laparotomy for Question 6. Because previous work has shown that administrative data may lack sufficient clinical detail to compare outcomes,[26] we did not attempt to directly compare complication rates between these procedures, or between diagnoses.

The search was limited to women 15 years and older, who had one of the following *International Classification of Diseases, Ninth Revision* (ICD-9) diagnostic codes: 183.x (malignant neoplasm of the ovary and other uterine adnexa), 220.x (benign neoplasms of the ovary); 620.x (ovarian cysts); 752.11 (para-ovarian cysts); 614.0, 614.1, 614.2, 614.6 (adnexal masses secondary to pelvic inflammatory disease); 789.33, 789.34, 789.39 (abdominal masses arising in the left or right lower quadrant, or other nonspecified site); and V655 (normal findings after diagnostic evaluation).

In order to avoid overestimation of complication rates due to other procedures, we then excluded patients who had an ICD-9 diagnosis code for hysterectomy (68.x). Procedures were then classified as laparoscopy only (54.21), laparoscopy with conservative ovarian surgery (65.3x, 65.4x, 65.5x, 65.6x), laparoscopy with oophorectomy (65.0x, 65.2x), or laparotomy (54.11) alone, with conservative ovarian surgery (same codes), or with oophorectomy (same codes).

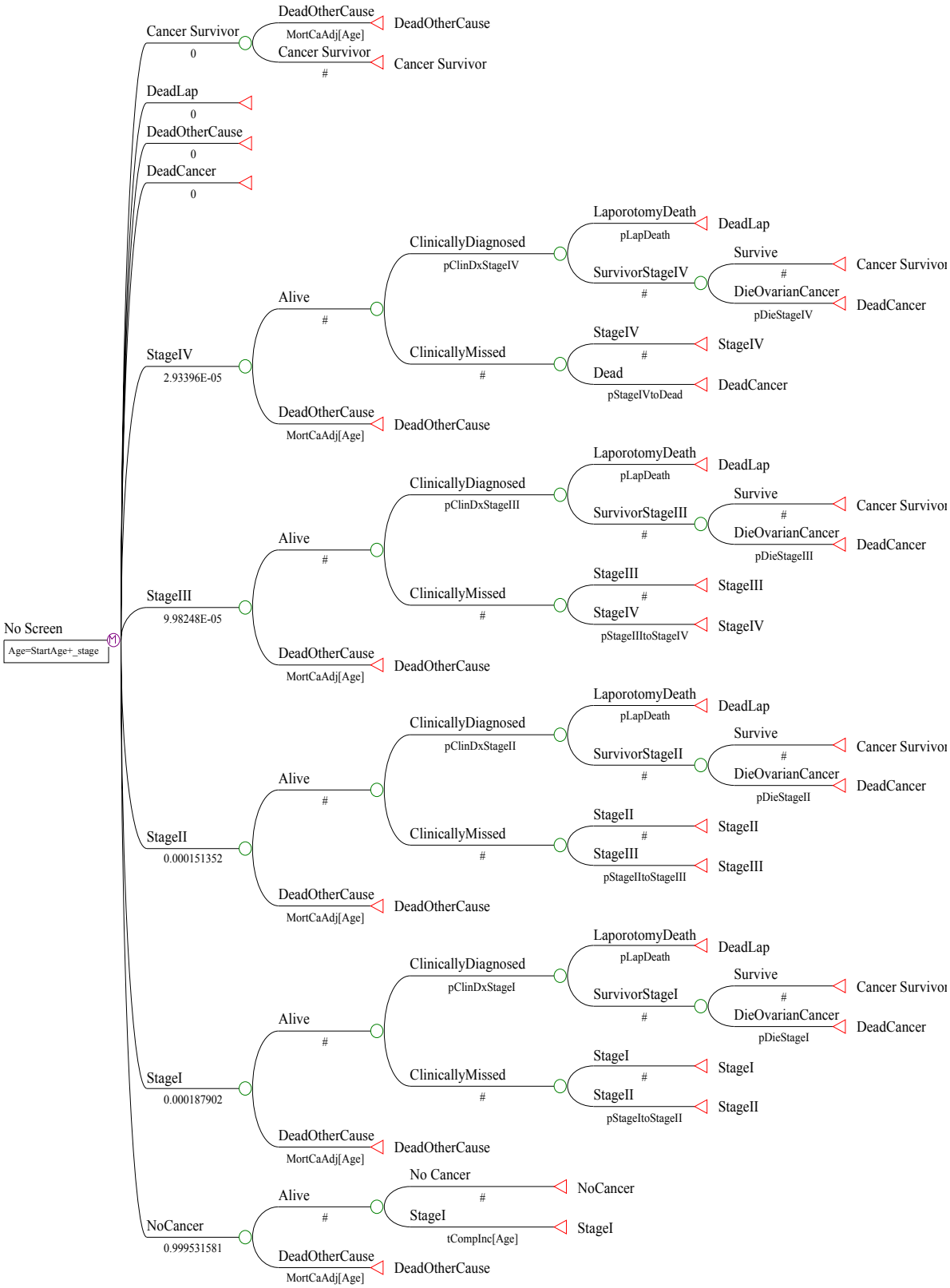
A discharge status of "Dead" indicated in-hospital mortality. Complications of surgery or hospitalization were indicated by diagnosis codes of E870 through E876.

Model of Natural History of Ovarian Cancer

We used a Markov state-transition model to explore the impact of alternate assumptions about the natural history of ovarian cancer. The original model was developed as a graduate school project by Karen Hoffman, MD, and further refined in collaboration with two of the authors of this report (Drs. Kulasingam and Myers).

The model simulates a cohort beginning at age 40 distributed across cancer stages. Subjects progress from no cancer through the stages of ovarian cancer to death. Each cycle is 12 months long. The original model design is illustrated in Figure 3; subsequent modifications include removal from the at-risk population by undergoing oophorectomy for another cause, and allowing some Stage I cancers the possibility of progressing directly to Stage III. Model variables and the ranges over which they were varied are outlined in Table 4. Probability of progressing from no cancer to Stage I cancer varies by age and is based on age-adjusted ovarian cancer incidence rates. Because the probability of progression (or duration of time within a stage) is unknown, probability of progression from Stage I to II, from Stage II to III, and from Stage III to Stage IV was adjusted to reflect incidence distribution across stages. Within the model, subjects may die from causes other than ovarian cancer. The probability of dying from a cause other than ovarian cancer varies by age and was constructed from CDC National Vital Statistics reports and Surveillance, Epidemiology, and End Results (SEER) data.[27,28] Probability of clinical diagnosis is based on the annual report of the International Federation of Gynecology and Obstetrics (FIGO).[29] Five-year survival rates gathered by SEER 1992-98 were used to predict probability of dying from ovarian cancer.[27] SEER localized disease corresponds to Stage I cancer, regional disease corresponds to Stage II cancer, and distant disease corresponds to Stage III/IV ovarian cancer. The model was constructed in DATA 4.0.[30]

Figure 3. Schematic of natural history model



Abbreviations for probabilities are described in Table 4, below.

Table 4. Model variables

Variable description	Model abbreviation of variable	Value	Range varied
Probability of clinical diagnosis for each stage (I, II, III, or IV) if no screening test or if screening produces a false negative	pClinDxStageI	0.261	Calibrated
	pClinDxStageII	0.446	
	pClinDxStageIII	0.837	
	pClinDxStageIV	0.950	
Probability of dying from diagnostic exploratory laporotomy	pLapDeath	0.00023	0.00 to 0.0010
Probability of dying from each stage of cancer, based on 5-year survival rates	pDieStageI	0.051	Not varied
	pDieStageII	0.187	Not varied
	pDieStageIII	0.691	Not varied
	pDieStageIV	0.691	Not varied
Probability of developing Stage I cancer, based on ovarian cancer incidence rates	tCompInc		Varies with age
Probability of dying from a cause other than ovarian cancer	tMortCaAdj		Varies with age

Peer Review Process

We employed internal and external quality-monitoring checks through every phase of the study to reduce bias, enhance consistency, and verify accuracy. Examples of internal monitoring procedures include: three progressively stricter screening opportunities for each article (abstract screening, full-text article review, data abstraction review); involvement of three individuals (two clinicians and copy editor) in each data abstraction; agreement of at least two clinicians on all included studies.

Our principle external quality-monitoring device was the peer-review process. Nominations for peer reviewers were solicited from several sources, including a technical expert panel and interested federal agencies. The list of nominees was forwarded to the Agency for Healthcare Research and Quality (AHRQ) for vetting and approval. A final list of peer reviewers is provided in Appendix E.*

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

Chapter 3. Results

Question 1: Prevalence of Tumor Types

Question 1 is: What is the prevalence of various tumor types among women with an adnexal mass, stratified by cancer status (malignant vs. benign), age, menopausal status, and size of tumor?

Approach

We included studies in the U.S. population with more than 50 women and limited the literature search to screening studies and case series where results were provided for all women with an undiagnosed mass, not just those with subsequent positive additional tests.[21] Studies of adnexal mass in which the gold standard is applied only to those with positive tests results would underestimate the prevalence of disease and cause a substantial bias.

Results

Twenty articles met the inclusion criteria and are described in the Evidence Table 1 (Appendix D*).[31-50]

Detailed prevalences for specific tumor types are provided in Evidence Table 1. The included studies can be divided into two groups. The first group includes four reports from a large screening study in Kentucky (Table 5). The prevalence of malignancy ranged from 0.09 to 0.18 percent. In postmenopausal women, the prevalence of malignancy was 0.09 to 0.18 percent, borderline tumors were not reported, and the prevalence of benign tumors was 0.08 to 1.3 percent. In a population that included either postmenopausal women or those with a family history of breast, ovarian, or colorectal cancer, the prevalence of malignancy was 0.10 to 0.11 percent, of borderline tumors 0.02 percent, and of benign tumors 1.1 to 1.2 percent.

The most common malignant tumor types include primary ovarian carcinoma, such as serous and mucinous cystadenocarcinoma, granulosa cell tumors, and undifferentiated adenocarcinoma. Borderline tumors were less common, such as serous low malignant potential (0.02 percent). The most common benign tumors were serous cystadenoma (0.4 to 0.7 percent), paratubal cyst (0.1 to 0.16 percent), endometrioma (0.03 to 0.3 percent), and mature teratoma (0.02 to 0.08 percent).

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

Table 5. Prevalence of tumor types in screening studies*

Study	N	% Menopausal	Malignant	Borderline	Benign
DePriest et al.,1993[36]	3,220	100; most had positive family history of breast, ovarian, or colorectal cancer	0.09%	Not reported	1.3%
DePriest et al.,1997[34]	6,470	Either menopausal or had positive family history of breast (30%), ovarian (24%), or colorectal cancer (15%)	0.11%	Not reported	1.2%
Modesitt et al., 2003[40]	15,106	100	0.18%	Not reported	0.8%
Van Nagell et al., 2000[49]	14,469	Either menopausal or had positive family history of breast (34%), ovarian (23%), or colorectal cancer (23%)	0.1%	0.02%	1.1%

*Note: All four publications represent the same screening study at different times.

The majority of U.S. studies with histological diagnosis of all masses (n = 16) were case series of women with undiagnosed adnexal mass undergoing laparotomy (Table 6). The prevalence of malignancy ranged from 5.7 to 57.5 percent, the range of borderline tumors was 1.4 to 11.2 percent, and the prevalence of benign tumors was 40 to 100 percent. All tumor types were over-represented because patients had an undiagnosed adnexal mass, and the clinical presentation was not well described in the majority of studies. Most studies included both premenopausal women and postmenopausal women and did not provide results separately. The one study that included only postmenopausal women[41] found only benign tumors.

Table 6. Case series and retrospective medical record reviews

Study	Denominator	Location	Age, menopausal status, race	Malignant	Borderline	Benign
Childers et al.,1996[32]	138	AZ	52	13.8%	Not reported	86.2%
Dottino et al.,1999[37]	160	NY	52.2 53% post 91% white	8.1%	5%	86.9%
Fleischer et al., 1996[38]	62	TN	50 >50% post	50%	Not reported	50%
Lin et al., 1993[39]	80	NY	56 76% post 90% white	57.5%	2.5%	40%
Parker et al., 1994[41]	61	Multi-site	65 100% post	None	None	100%
Roman et al., 1997[42]	226	CA	20% post	11.5%	7.5%	81%
Schneider et al., 1993[43]	55	AZ	53 60% post	25.5%	3.6%	70.9%

Study	Denominator	Location	Age, menopausal status, race	Malignant	Borderline	Benign
Scoutt et al., 1994[44]	109	CT	40	20.2%	Not reported	79.8%
Shen-Gunther et al., 2002[45]	125	OK/NV	58 82% white	44.8%	9.6%	45.6%
Smikle et al., 1995[46]	195	TX	40% post	13.3%	Not reported	86.7%
*Chalas et al., 1992[31]	241	NY	Not reported	50.2%	7.5%	42.3%
Cohen et al., 2001[33]	71	IL	22-80 44% post	18.3%	1.4%	80.3%
DePriest et al., 1993[35]	121	KY	3-74 49% post	10.7%	Not reported	89.3%
Troiano, 1997[47]	144	CT	45 29% post	11.8%	2.1%	86.1%
Twickler et al., 1999[48]	244	TX	38.6	5.7%	6.6%	87.7%
Vasilev et al., 1988[50]	182	CA	Not reported	8.2%	1.6%	90.1%

*Retrospective chart review

Discussion

Estimating the age-specific prevalence of specific adnexal tumor types from the available literature is difficult. The best data come from a series of reports from a large screening study; overall prevalence of masses was 1 to 2 percent, with benign masses outnumbering malignant by 4- to 10-fold. Because patients with negative screening test results did not undergo definitive diagnostic procedures in these studies, the prevalence estimates are dependent on the sensitivity of the screening tests used (and the completeness of followup among test negatives). In addition, there is a potential bias in that premenopausal women enrolling in the screening study were at higher risk than average because of family history; in addition, postmenopausal women may have been more likely to enroll because of concerns based on family history, vague symptoms, or other reasons which would affect relative prevalence compared to the general population.

Estimates of prevalence in studies with 100 percent histologic diagnosis are inevitably biased by the clinical factors that determine which patients ultimately undergo surgery. These can include the presence and nature of symptoms (patients with symptoms referable to a mass would likely undergo surgery sooner than those with asymptomatic masses, all other things being equal); other findings (for example, the presence of ascites); patient anxiety; the diagnostic algorithms used (for example, the duration of followup for persistence); and the nature of the practice (malignancies will be more frequent in a gynecologic oncology practice compared to a general gynecology practice).

As mentioned previously, we did not include studies from outside the United States. Given differences in ethnic backgrounds (affecting genetic risks), observed differences in cancer incidence, and differences in clinical practice between countries, and the almost universal failure of studies to describe the clinical history leading to the diagnosis of adnexal mass, inclusion of these studies would not have allowed a more precise estimate of prevalence of different types of adnexal masses in the U.S. population.

Summary

In four reports from a large U.S. screening study, the prevalence of adnexal masses detected by ultrasound among postmenopausal women was 0.8 to 1.3 percent, and the prevalence of malignancy 0.09 to 0.18 percent (i.e., 9 to 18 per 10,000). Prevalence of different pathologies varies widely among case series. There are no data on the relative prevalence of different pathologies among women with asymptomatic masses compared to women with symptomatic masses.

Question 2: Bimanual Pelvic Examination

Question 2 is: What are the sensitivity, specificity, and reliability of the bimanual pelvic examination?

Approach

Articles were sought which evaluated the ability of the bimanual examination to detect adnexal masses, and/or to discriminate benign from malignant masses. Preference was given to studies where there was histological confirmation of the diagnosis, but an alternative reference standard (such as followup) was allowed for screening studies. Data allowing calculation of sensitivity and specificity had to be provided.

Our rationale for including the pelvic examination was based on its role in the initial evaluation of adnexal masses. Some asymptomatic women will have a mass detected as part of a "routine" physical examination; others will have a mass detected as part of an examination performed because of symptoms. The postexamination probability of malignancy is a function of the prevalence of cancer and the sensitivity and specificity of the bimanual examination; these probabilities, in turn, will affect the positive and negative predictive values of additional tests such as cancer antigen 125 (CA-125) and imaging studies. Because the pelvic examination will be the first test performed, either as a screening test or as a diagnostic test, knowledge of its test characteristics is important for evaluating subsequent diagnostic tests.

Results of Literature Search and Screening

We identified 14 studies that met the inclusion criteria.[42,51-63] Nine studies provided data on discrimination between benign and malignant masses,[42,51,53-57,62,63], four on the ability of the bimanual examination to detect any adnexal mass,[52,59-61] and one provided data on both discrimination between benign and malignant and ability to detect masses.[58] All 14 studies are summarized in Evidence Table 2 (Appendix D*).

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

Study Characteristics

Types of data incorporated. Two of the studies[54,56] included history or clinical impression as part of the "test;" results were not provided separately for examination alone.

Types of study population. Ten of the 14 studies were performed prior to surgery for an adnexal mass, while four were from screening studies.[51,52,57,58]

Reporting of study populations. Of the screening studies, Andolf et al.[52] was performed in women over 40 considered at high risk of ovarian cancer because of symptoms or risk factors; Grover and Quinn[57] was performed in asymptomatic volunteers 25 and older, but described menopausal status; Adonakis et al.[51] was performed in women over 45; and Jacobs et al.[58] was done entirely in a postmenopausal population.

Seven of the 11 preoperative studies reported menopausal status, but only two reported on test characteristics specifically by menopausal status.[55,56] None reported race/ethnicity, and none reported the clinical route by which patients had come to surgery (detection of an asymptomatic mass, symptoms, etc.).

Methodology. The methodological quality of the included studies was as follows:

Reference standard. Of the preoperative studies, all but one[42] had operative confirmation of findings. Ultrasound was used as the reference standard in the four screening studies, with 12-month followup examinations or questionnaires.

Verification bias. In the study by Roman et al.,[42] 26 women with non-palpable masses did not undergo definitive diagnosis.

Test reliability. Only one study[60] provided direct data on test reliability. Grover and Quinn,[57] Ong et al.,[59] Schutter et al.,[63] and Buckshee et al.[54] used a single examiner. The other studies did not address the issue of test reliability.

Sample size. None of the reports had a priori sample size calculations.

Use of appropriate statistical tests. All reports used appropriate techniques for calculating test characteristics.

Blinding. Only two studies[54,60] explicitly stated whether examiners were blinded to prior history or other findings.

Definition of positive and negative test. Nine of 14 studies reported their definitions of a positive test, although the precision of the definitions was quite variable (from "a mass 5 cm or more in diameter" to "larger than normal"); others relied on "clinical impression."

Results

Table 7 and Figure 4 present the results of studies that evaluated the sensitivity of the bimanual examination for detecting an adnexal mass. The studies of Padilla et al.[60,61] are particularly striking for the low sensitivity, since the examinations were performed under anesthesia, when, presumably, patient discomfort would not be a limiting factor. Both studies suggested a relationship between experience and accuracy; medical students performed worse than residents, who performed worse than attending physicians. Although these differences were not statistically significant, the studies were underpowered to detect significant differences. Obesity, defined as a body mass index greater than 30, had a significant negative impact on sensitivity, while increasing uterine size increased sensitivity, possibly by elevating the adnexae out of the pelvis.

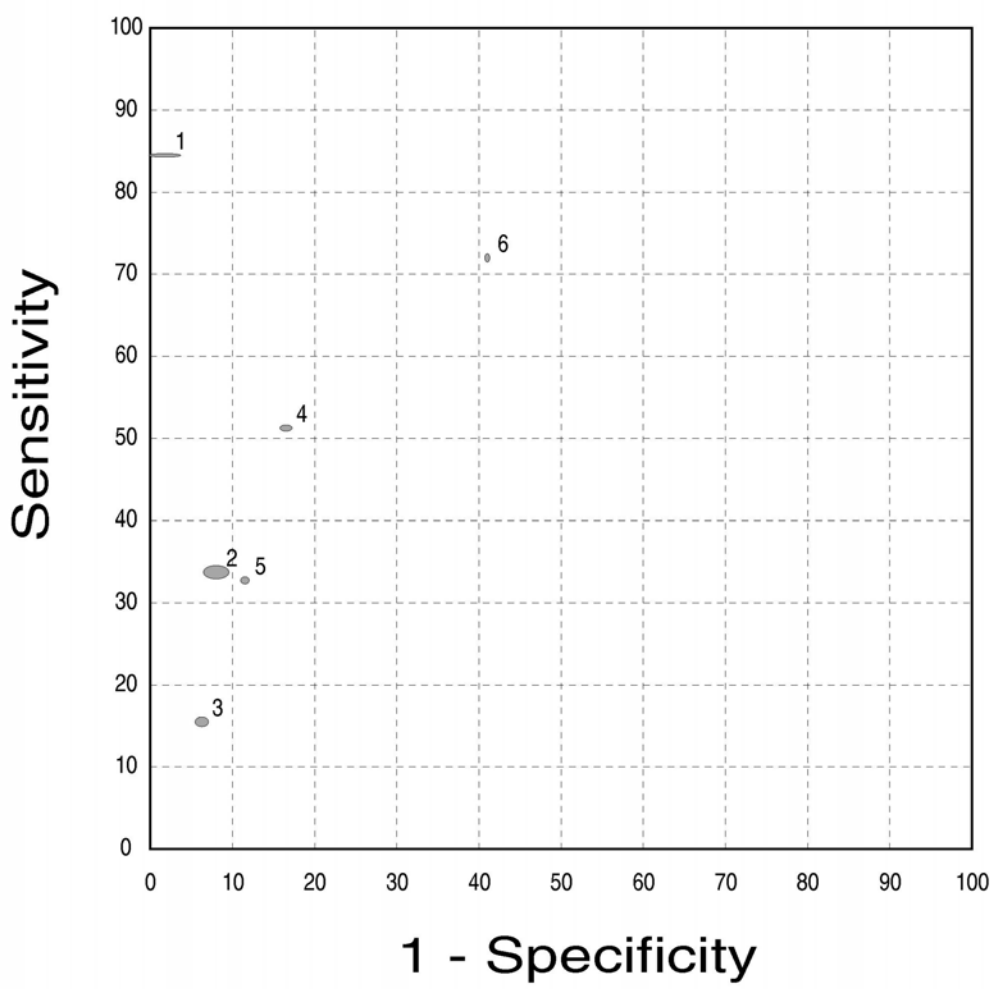
When sensitivity and specificity were combined separately using a random-effects model, the pooled sensitivity was 0.45 (95% confidence interval [CI], 0.28 to 0.68), and the pooled specificity was 0.90 (0.80 to 0.96).

Table 7. Sensitivity and specificity of pelvic examination in detecting the presence of an adnexal mass

Study	N	Sensitivity (95% CI)	Specificity (95% CI)	% with confirmed mass	Notes
Jacobs et al., 1988[58]	1,010	84.6% (65.0 to 100%)	98.3% (97.5 to 99.1%)	1.3% (0.1% malignant)	Reference standard: ultrasound Screening study
Andolf et al., 1990[52]	801	33.7% (26.5 to 41.0%)	92.0% (89.9 to 94.1%)	20% (0.1% malignant)	Reference standard: ultrasound by midwife Screening in women considered at high risk for ovarian cancer; no ovarian cancers detected: 2 endometrial cancers, 1 LMP detected
Padilla et al., 2005[61]	252	15.6% (8.1 to 23.0%)	93.8% (90.1 to 97.5%)	35.7% (unclear if any malignancies)	Exam under anesthesia prior to surgery for pelvic mass; examiners blinded to radiology findings Likelihood of not detecting an adnexal mass increased with less experience (OR for resident 1.13, student 1.36 compared to attending, although 95% CIs cross 1). Statistically significant increase in missed diagnosis if subject with BMI > 30 (OR 2.57; 95% CI, 1.36 to 4.87), and significant decrease in presence of enlarged uterus (OR 0.48; 95% CI, 0.25 to 0.93). Final diagnoses not presented, reasons for surgery not systematically presented
Padilla et al., 2000[60]	140 (82 masses)	Left adnexa (attending exam): 32.7% (19.5 to 45.8%) Right adnexa (attending exam): 21.2% (7.3 to 35.2%)	Left adnexa (attending exam): 88.5% (81.4 to 95.6%) Right adnexa (attending exam): 78.7% (70.4 to 87.0%)	58% (0 malignancies)	Exam under anesthesia prior to surgery for pelvic mass; examiners blinded to radiology findings; no clear relationship to experience
Ong et al., 1996[59]	86	71.9% (60.9 to 82.9%)	59.1% (38.5 to 78.6%)	74.4% (0 malignant)	Pre-surgical exam

Abbreviations: BMI = body mass index; CI = confidence interval; LMP = low malignant potential tumor; OR = odds ratio

Figure 4. Performance of bimanual pelvic examination for detecting the presence of an adnexal mass



Key to Figure 4: 1 = Jacobs et al., 1988;[58] 2 = Andolf et al., 1990;[52] 3 = Padilla et al., 2005;[61] 4 = Roman et al., 1997;[42] 5 = Padilla et al., 2000;[60] 6 = Ong et al., 1996[59]

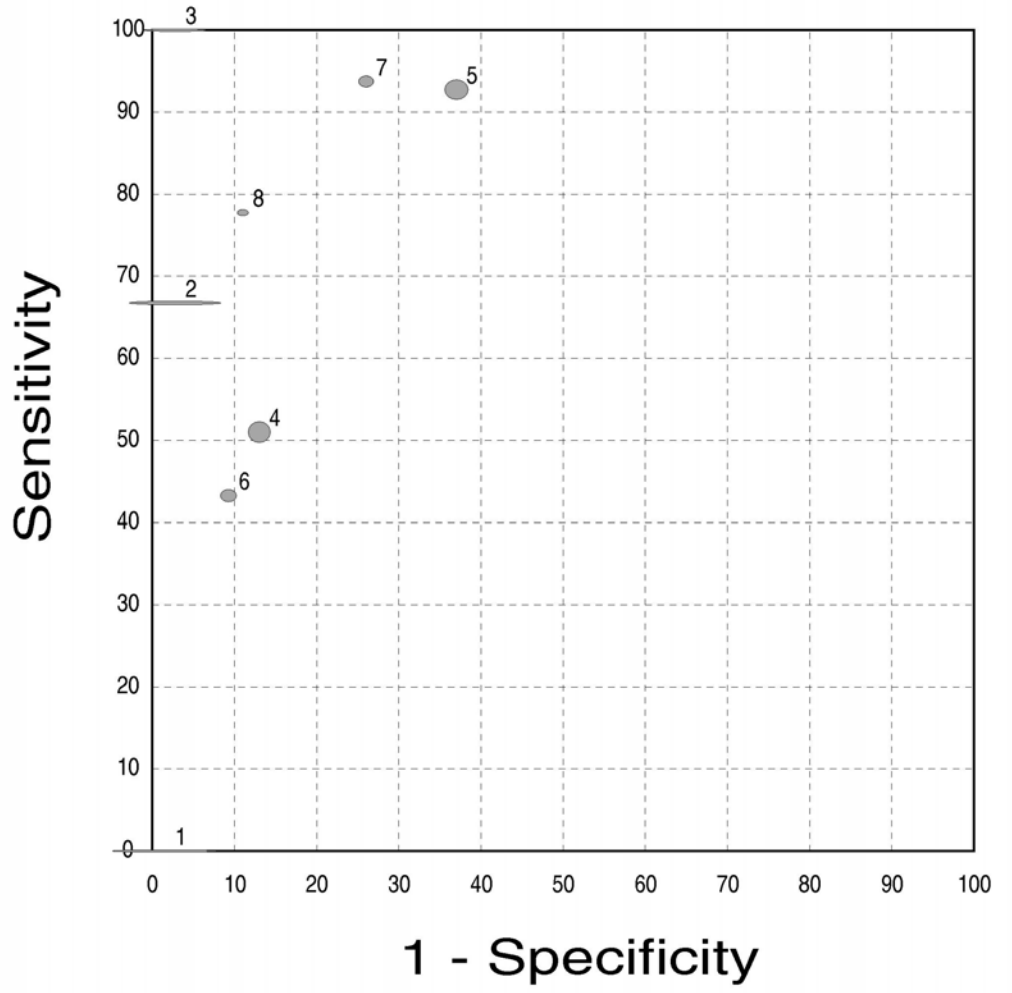
Table 8 and Figure 5 show the test characteristics for discriminating benign from malignant masses. Using a random-effects model, pooled sensitivity was 0.72 (95% CI, 0.49 to 0.88) and specificity was 0.92 (0.80 to 0.97). When only the three screening studies were included, pooled sensitivity was 0.58 (95% CI, 0.21 to 0.88), pooled specificity 0.98 (0.97 to 0.98).

Table 8. Sensitivity and specificity of pelvic examination in discriminating benign from malignant adnexal masses

Study	N	Sensitivity (95% CI)	Specificity (95% CI)	% Malignant	Notes
Adonakis et al., 1996[51]	2,000	66.7% (13.3 to 100%)	97.2% (96.5 to 97.9%)	0.15%	Screening study; threshold of "abnormal or ambiguous exam;" CA-125 used in conjunction to proceed to ultrasound
Grover et al., 1995[57]	2,623	0% (0 to 100%)	98.5% (98.0 to 98.9%)	0.05%	Screening study; ultrasound and clinical followup
Jacobs et al., 1988[58]	1,010	100% (0 to 100%)	97.3% (96.3 to 98.3%)	0.1%	Screening study; followup with ultrasound
Roman et al., 1997[42]	200	51.2% (36.3 to 66.1%)	83.6% (77.8 to 89.4%)	21%	Results for 26 patients with non-palpable masses not included; no substantial difference based on menopausal status
Buckshee et al., 1998[54]	34	77.8 % (50.6 to 100%)	88.9% (77.0 to 100%)	25%	One examiner; non-consecutive patients prior to surgery
Balbi et al., 2001[53]	72	90% (77.5 to 100%)	74% (61.8 to 86.2%)	31%	18 patients with "clearly benign masses" and 2 with "clearly malignant" excluded; clinical impression
Finkler et al., 1988[56]	106	43.2% (27.3 to 59.2%) Premenopausal: 16.7% (0 to 33.9%) Postmenopausal: 68.4% (47.5 to 89.3%)	90.8% (83.7 to 97.8%) Premenopausal: 92.3% (85.1 to 99.6%) Postmenopausal: 84.6% (65.0 to 100%)	36% Premenopausal: 26% Postmenopausal: 59%	"Clinical impression" included exam plus history; results not calculated for exam alone
Schutter et al., 1998[63]	155	91.5% (84.4 to 98.6%)	73.9% (64.9 to 82.9%)	39%	All postmenopausal; high prevalence of cancer; single examiner; inclusion/exclusion criteria not described
Schutter et al., 1994[62]	222	92.6% (87.4 to 97.9%)	63.0% (54.6 to 71.4%)	43%	Preoperative patients
Dowd et al., 1993[55]	225	51.0% (41.7 to 60.3%) Premenopausal: 31% Postmenopausal 59%	87.0% (80.8 to 93.2%) Premenopausal: 95% Postmenopausal: 75%	49%	Preoperative patients

Abbreviations: CA-125 = cancer antigen 125; CI = confidence interval

Figure 5. Performance of bimanual pelvic exam for distinguishing benign from malignant adnexal masses



Key to Figure 5: 1 = Grover and Quinn, 1995;[57] 2 = Adonakis et al., 1996;[51] 3 = Jacobs et al., 1988;[58] 4 = Dowd et al., 1993;[55] 5 = Schutter et al., 1994;[62] 6 = Finkler et al., 1988;[56] 7 = Balbi et al., 2001;[53] 8 = Buckshee et al., 1998[54]

For both types of studies, there appears to be a trend towards decreased specificity as prevalence increases, although the number of studies is small and the confidence intervals are wide. The extreme differences in sensitivity in the two largest studies (0 and 100 percent) prevent even a qualitative assessment of any relationship between prevalence and sensitivity.

The two studies that stratified results by menopausal status[55,56] found lower sensitivity and higher specificity for discriminating benign from malignant masses in premenopausal women compared to postmenopausal women (Table 8).

Discussion

Despite the common recommendation for routine pelvic examination, we found surprisingly little literature on its accuracy. Based on the literature we did identify, its sensitivity for detecting adnexal masses appears fairly low. Sensitivity for detecting normal adnexa is also low, as demonstrated in a recent study of examinations under anesthesia.[64] Although sensitivity for

distinguishing a malignant mass from a benign one is somewhat better, these results need to be interpreted with caution, since most of the studies were done in preoperative patients, who would already have a higher probability of having a malignancy. In the four large screening studies, there was a total of only five malignancies, with the bimanual detecting 0 percent, 66 percent, and 100 percent in the three individual studies where ovarian cancer was detected; the fourth had one case of a low malignant potential tumor and two endometrial cancers. Pooled sensitivity for the three screening studies that addressed discrimination between benign and malignant masses was considerably lower than for all studies combined (and was similar to the pooled sensitivity of the studies that examined the ability to detect any adnexal mass).

Both types of studies show a trend toward decreased specificity as the prevalence of abnormality increases – this may reflect a greater degree of suspicion on the part of the examiner, based on other findings, and a greater likelihood of calling an examination abnormal. This is supported by the finding of the two studies which stratified results by menopausal status, which found higher sensitivity and lower specificity in postmenopausal women compared to premenopausal women.[55,56] Because examiners were unblinded, and were likely aware of the higher prevalence of malignancy among postmenopausal women, they may have been more likely to assign a diagnosis of malignancy among those patients. Future studies need to pay stricter attention to blinding examiners to other information. In theory, this bias should also result in higher sensitivity as prevalence increases, although, because of the small number of studies, the small numbers of subjects in most studies, and the diametrically opposed findings of the two largest studies, we were unable to recognize any relationship.

In the two studies that addressed the effect of experience on test characteristics,[60,61] there appeared to be a relationship between increasing experience and increased sensitivity (specificity did not change); however, even attending physicians achieved a sensitivity of only 28 percent. Based upon the available literature, the bimanual examination does not appear to be a sensitive test for detecting the presence of adnexal masses and appears to have limited ability to discriminate benign from malignant masses. Although specificity was somewhat better, positive predictive values will still be quite low in low prevalence settings, as discussed under Question 7. This will, in turn, lower the positive predictive value of diagnostic tests performed in patients referred on the basis of a pelvic examination. These tests are discussed in detail in the next section.

Question 3: Single Modality Tests

Question 3 is: Among women with a palpable adnexal mass on exam or a mass identified by ultrasound/imaging, what is the sensitivity/specificity of various evaluation modalities including ultrasound (transvaginal ultrasound [TVUS], transabdominal ultrasound, color Doppler, two-dimensional [2D] versus three-dimensional [3D] ultrasound), computer tomography (CT) scan, magnetic resonance imaging (MRI) scan, and CA-125 levels for distinguishing benign from malignant masses?

Approach

This section considers the various evaluation modalities that are described in the literature and would be available to a clinician to aid in the work-up of an adnexal mass after it has been

diagnosed. We focused our search on articles whose primary reference standard was histopathology. Ideally this reference standard would be applied to all test negatives. However, we accepted a repeat negative test (such as imaging) conducted at least 6 months later as an acceptable alternative. We did include some studies that were from population-based screening samples, and these will be considered in a separate section below. The evaluation modalities investigated can be divided into several general categories. Imaging studies will be divided by technological mode (ultrasound, MRI, etc.). Ultrasound studies will be divided into those that evaluate adnexal morphology (either by an explicit scoring system or by descriptive standards), those that measure vascular flow in the mass (Doppler), and those that evaluate these modalities in combination. Serum studies will focus primarily on CA-125, as this is the most common marker in both the literature and in clinical practice. However, other serum markers will be discussed as well. Finally, the studies for which it was possible to stratify by menopausal status will be discussed where appropriate.

Results of Literature Search and Screening

Two hundred and five articles were identified for abstraction. Of these, 153 met the inclusion criteria and were abstracted into Evidence Table 2 (Appendix D*).[31,33-36,39,42-44,46,47,49-56,58,62,63,65-195]

Ultrasound Morphology

Conventional grey scale ultrasonography is the most common imaging modality used to differentiate benign from malignant adnexal masses. Especially with the advent of high-frequency transvaginal probes, the quality of the images allows description of the gross anatomic features of the lesion. This is, however, limited by the great variability of macroscopic characteristics of both benign and malignant masses. Furthermore, the technique is operator dependent. To overcome these limitations, morphologic scoring systems have been developed. Such scoring systems are based on specific ultrasound parameters each with several scores according to determined features and with a cutoff value to categorize masses as either malignant or benign.

Table 9 describes the details of the most commonly used scoring systems. Briefly, the following scores are suggestive of malignancy: Sassone[159] greater than 9, DePriest[36] greater than or equal to 5, Ferrazzi[93] greater than 9, and for Lerner[131] greater than or equal to 3. Although the development of all the scoring systems was motivated to improve the reproducibility of morphological measurements, only the scoring system by Lerner et al. based the categories on a multivariate logistic analysis.

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

Table 9. Detailed description of ultrasound scoring systems

Scoring system	Score				
Sassone et al., 1991[159]					
Morphology	1	2	3	4	5
Inner wall structure	Smooth	Irregularities ≤ 3 mm	Papillarities > 3 mm	Not applicable, mostly solid	-
Wall thickness (mm)	Thin (≤ 3)	Thick (> 3)	Not applicable, mostly solid	-	-
Septa (mm)	None	Thin (≤ 3)	Thick (> 3)	-	
Echogenicity	Sonolucent	Low echogenicity	Low echogenicity with ochogenic core; mixed echogenicity	-	High echogenicity
DePriest et al., 1993[36]					
Morphology	0	1	2	3	4
Cystic wall structure	Smooth (< 3 mm thick)	Smooth (> 3 mm thick)	Papillary projection (< 3 mm)	Papillary projection (≥ 3 mm)	Predominately solid
Volume (cm3)	< 10	10-50	> 50-200	> 200-500	> 500
Septum structure	No septa	Thin septa (< 3 mm)	Thick septa (3 mm to 1 cm)	Solid area (≥ 1 cm)	Predominately solid
Ferrazzi et al., 1997[93]					
Morphology	1	2	3	4	5
Wall	≤ 3 mm	> 3 mm	-	Irregular, mostly solid	Irregular, not applicable
Septa	None	≤ 3 mm	> 3 mm		
Vegetations	None	-	-	≤ 3 mm	> 3 mm
Echogenicity	Sonolucent	Low echogenicity	-	With echogenic areas	With heterogeneous echogenic areas, solid
Lerner et al., 1994[131]					
Morphology	0	1	2	3	
Wall structure	Smooth or small irregularities < 3 mm	-	Solid or not applicable	Papillarities ≥ 3 mm	
Shadowing	Yes	No	-	-	
Septa	None or thin (< 3 mm)	Thick (≥ 3 mm)	-	-	
Echogenicity	Sonolucent or low-level echo or echogenic core	-	-	Mixed or high	

Reproducibility of tests. Timmerman et al.[196] evaluated the subjective assessment of ultrasonographic images for discriminating between malignant and benign masses. Three hundred consecutive patients were evaluated with TVUS by six different operators, and both

diagnostic accuracy and interassessor agreement were calculated. The operators had varied experience in TVUS – from approximately 300 to 15,000 scans. The two most experienced operators agreed 92 percent of the time. The accuracy of the least experienced operators ranged from 82 to 87 percent ($p = 0.0001$). Overall, 65 percent of all the masses were correctly classified by all six operators. Interassessor agreement was greater between the most experienced operators as well (kappa = 0.852). When comparing experienced with less experienced operators, the kappa ranged from 0.581 to 0.737. This is similar to the kappa reported by Yamashita et al.[192] among five operators, 0.62 (± 0.02) with TVUS. Interassessor agreement was not calculated between the less experienced operators. None of the included articles described operator experience, and only a few addressed interobserver variability. Although operator experience appears to correlate with accuracy, the specialty training of the unltrasonographer does not. In a meta-analysis of both morphologic and color Doppler tests in the evaluation of adnexal masses, Kinkel et al.[197] found no difference between radiologists and gynecologists in the performance of ultrasound.

TVUS versus abdominal ultrasound. Of the 122 articles that evaluated adnexal masses via ultrasound (through either ultrasound morphology or Doppler measurements), only five articles exclusively used transabdominal imaging.[52,58,116,133,198] Fifty-nine articles used TVUS exclusively and 51 used a combination of TVUS and abdominal ultrasound. There were seven articles for which the ultrasound modality was unknown. In the majority of the articles that used a combination of TVUS and abdominal ultrasound, TVUS was the "method of choice." The most common reasons cited for also including abdominal ultrasound were patient refusal of transvaginal scans, virginity, poor image quality, and very large masses. Although a few articles reported how many women had which type of ultrasound, none of the articles reported their results such as to permit a stratification by TVUS or abdominal ultrasound. We therefore elected to group all ultrasound studies together regardless of TVUS or abdominal imaging.

Trials identified. We identified 69 articles comprising 73 ultrasound morphology assessments. Despite the availability of published scoring systems, most of the studies based their diagnoses on either descriptive assessments of adnexal masses or used a modified or unique scoring system. Only 13 studies explicitly used Sassone's criteria, six used DePriest's, and three used Ferrazzi's, Finkler's, Lerner's, and Valentin's respectively. When a scoring system other than an established criterion was used, it was not always clear how it had been developed or modified. Details of the tests and their evaluative performance are provided in Table 10. Assessments of adnexal morphology by ultrasound which were either a unique or modified or unclear scoring system are labeled "other" with a brief description when possible. It is also important to note that not all of the established scoring systems were employed using the original cutpoints. For example, Caruso et al.[83] and Itakure et al.[115] both used a cutpoint of > 7 for the DePriest scoring system, where the original description used ≥ 5.

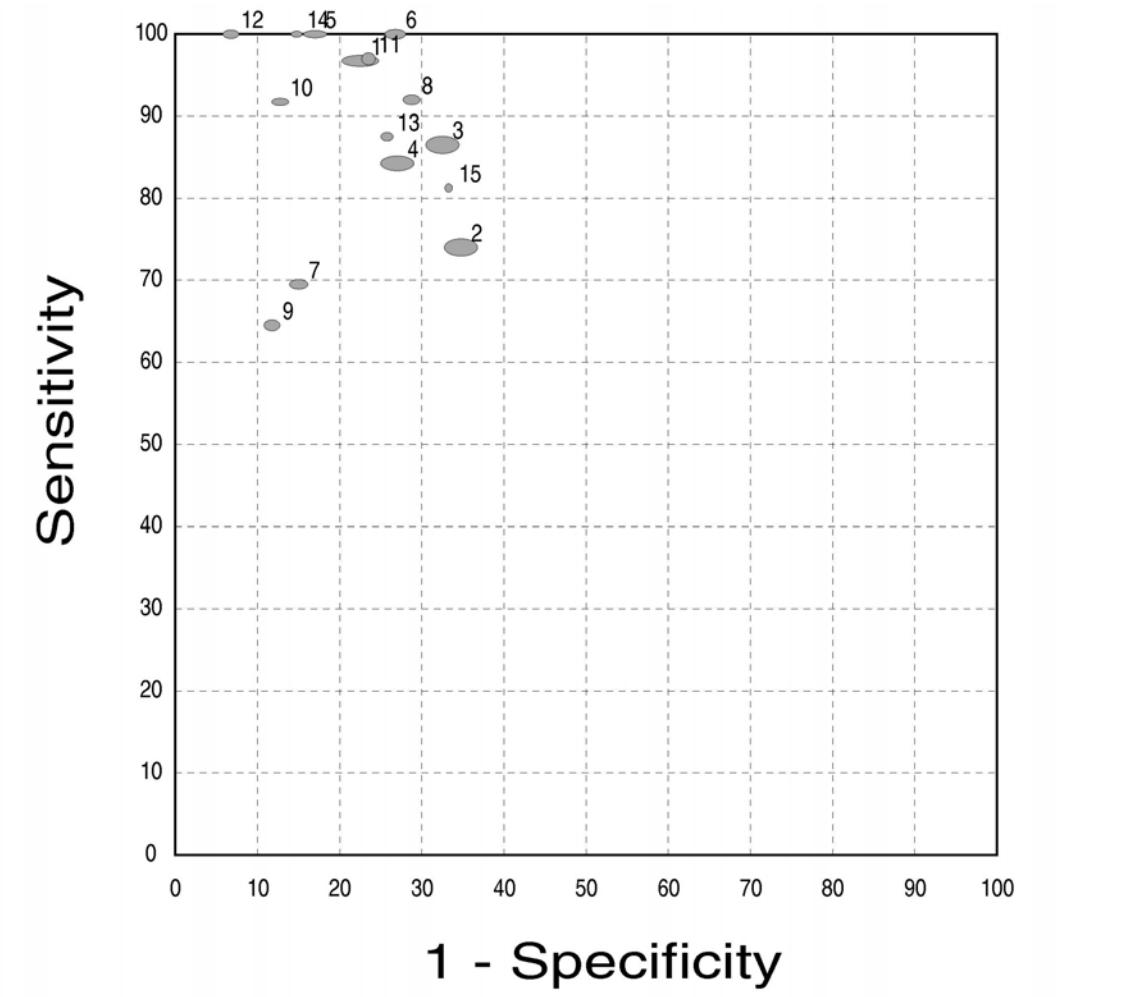
Table 10. Sensitivity and specificity of ultrasound morphology

Scoring system	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Range of sensitivity in individual studies	Range of specificity in individual studies	References
Sassone	0.86 (0.79 to 0.91)	0.77 (0.73 to 0.81)	0.65 to 1.00	0.65 to 0.93	43,54,68,69,83,93,130,131,154,159,160,163,179,193,199
DePriest	0.91 (0.84 to 0.95)	0.68 (0.49 to 0.82)	0.88 to 1.00	0.40 to 0.81	35,36,69,83,93,115
Ferrazzi	0.87 (0.80 to 0.92)	0.81 (0.62 to 0.91)	0.84 to 0.87	0.67 to 0.88	69,75,93
Finkler	0.82 (0.65 to 0.91)	0.78 (0.59 to 0.91)	0.52 to 0.88	0.55 to 0.70	56,62,63
Other (note: significant heterogeneity in criteria used for diagnosis – see ROC curve)	0.86 (0.82 to 0.89)	0.83 (0.76 to 0.88)	0.43 to 1.00	0.29 to 1.00	33,34,39,42,43,67,69,74,76-80,87,90,95,97,101,102,104,106,108,112,117,118,122,124-127,133-135,138-140,142,144,146,147,155,161,166,168,169,171,180,181,185,187,188,192,195

Abbreviations: CI = confidence interval; ROC = receiver operating characteristic

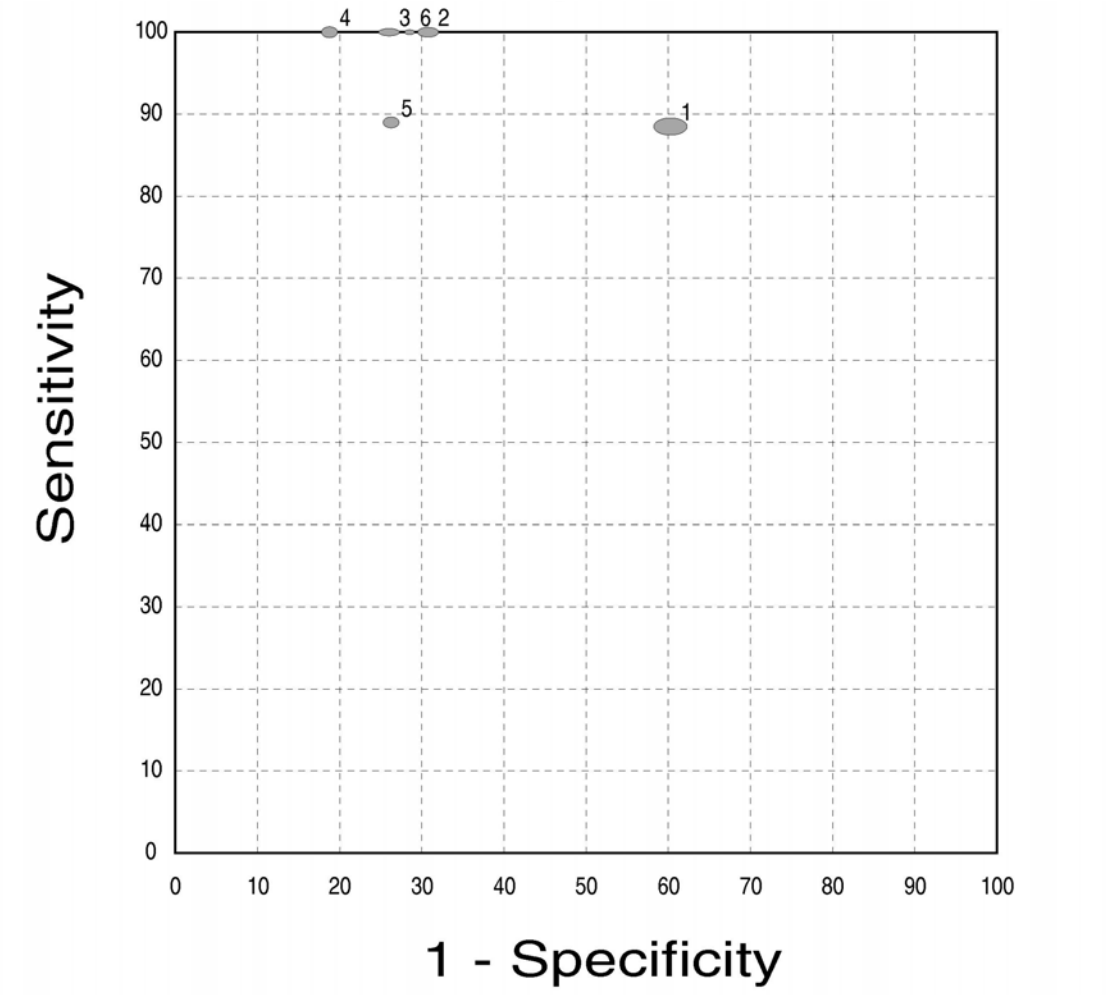
Results. The results of pooled sensitivity and specificity using a random-effects model, along with the range of sensitivity and specificity reported in individual studies, are shown in Table 10. Included studies are shown in Figures 6-10. There was a great range in test results, especially in the studies not using established scoring systems. This most likely reflects the heterogeneity of the tests themselves. There was little concrete difference among the established scoring systems. Overall the tests achieved relatively higher levels of sensitivity and negative predictive value (NPV) in the diagnosis of malignancy than specificity or positive predictive value (PPV). With the exception of four studies, the NPV was above 0.80, with the majority of tests above 0.90. The PPV in the majority of studies was below 0.50. In general, there was a trade-off between sensitivity and specificity, both in the individual studies of a specific scoring system, and in pooled results of all studies of a scoring system – as sensitivity increases, specificity decreases.

Figure 6. Performance of ultrasound scoring according to Sassone's criteria (1991)



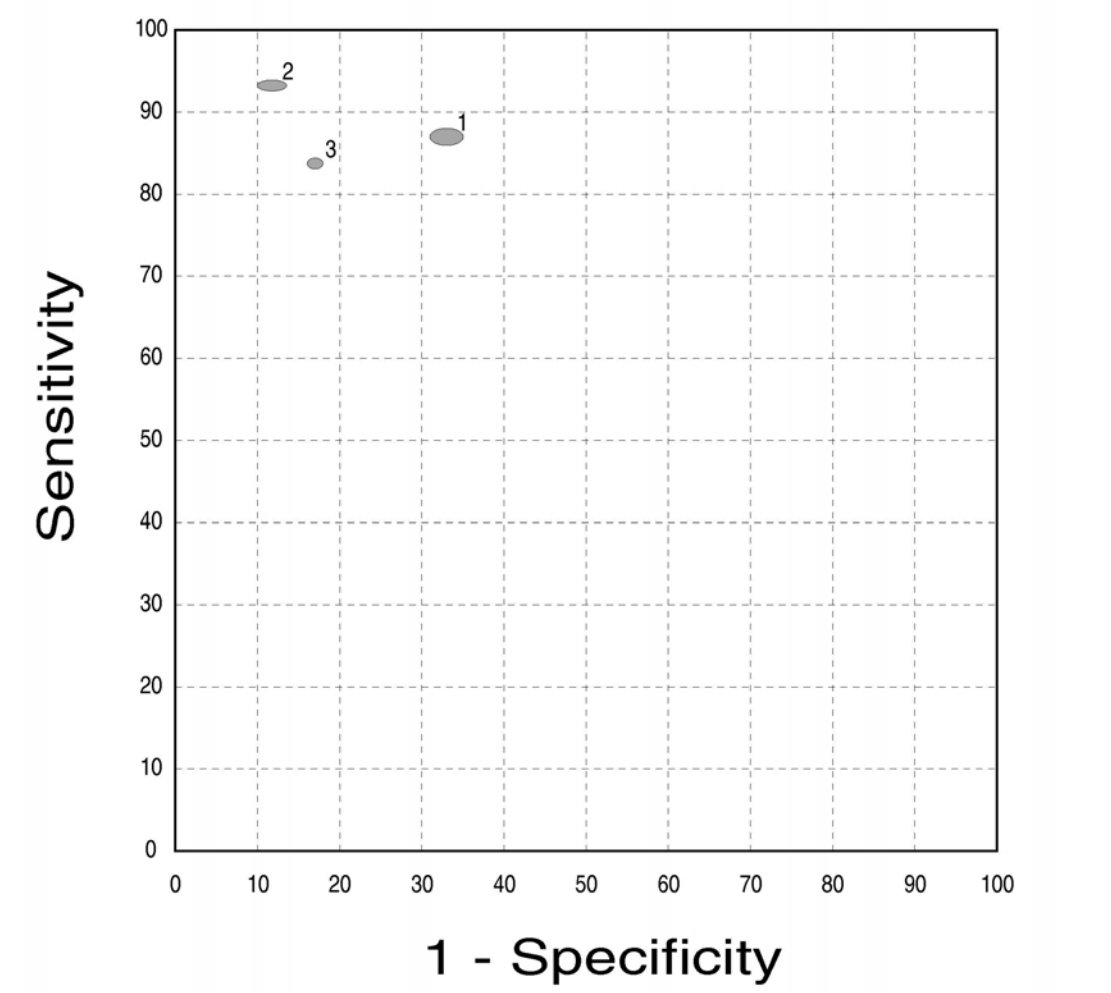
Key to Figure 6: 1 = Lerner et al., 1994;[131] 2 = Ferrazzi et al., 1997;[93] 3 = Sawicki et al., 2001;[160] 4 = Rehn et al., 1996;[154] 5 = Sassone et al., 1991;[159] 6 = Caruso et al., 1996;[83] 7 = Leeners et al., 1996;[130] 8 = Alcazar and Lopez-Garcia, 2001;[68] 9 = Alcazar et al., 2003;[69] 10 = Timor-Tritsch et al., 1993;[179] 11 = Zanetta et al., 1994;[193] 12 = Alcazar et al., 1996;[199] 13 = Schneider et al., 1993;[43] 14 = Buckshee et al., 1998;[54] 15 = Sengoku et al., 1994[163]

Figure 7. Performance of ultrasound scoring according to DePriest's criteria (1993)



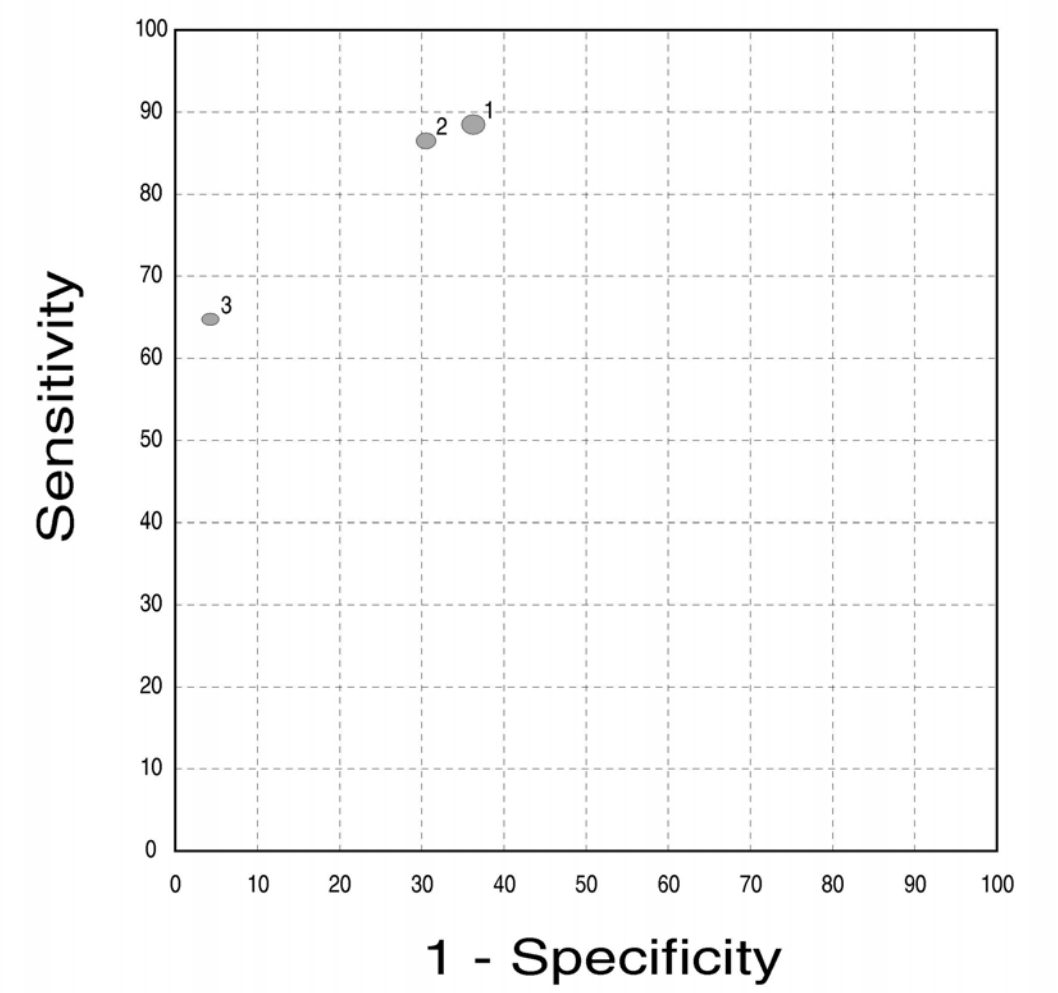
Key to Figure 7: 1 = Ferrazi et al., 1997;[93] 2 = Caruso et al., 1996;[83] 3 = DePriest et al., 1993;[35] 4 = Alcazar et al., 2003;[69] 5 = Itakura et al., 2003;[115] 6 = DePriest et al., 1993[36]

Figure 8. Performance of ultrasound scoring according to Ferrazzi's criteria (1997)



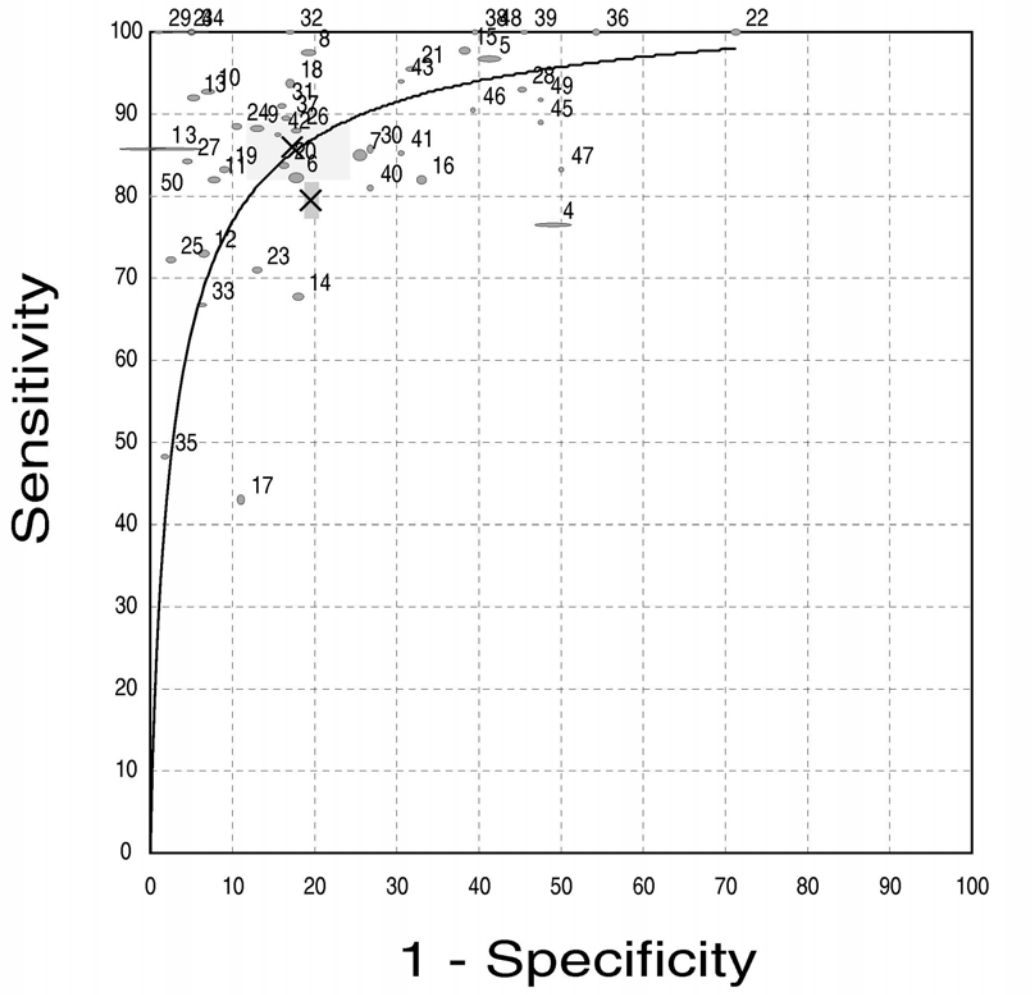
Key to Figure 8: 1 = Ferrazzi et al., 1997;[93] 2 = Berlanda et al., 2002;[75] 3 = Alcazar et al., 2003[69]

Figure 9. Performance of ultrasound scoring according to Finkler's criteria (1988)



Key to Figure 9: 1 = Schutter et al., 1994;[62] 2 = Schutter et al., 1998;[63] 3 = Finkler et al., 1988[56]

Figure 10. Performance of ultrasound scoring according to various other unvalidated criteria



Key to Figure 10: 1 = DePriest et al., 1997;[34] 2 = Marchetti et al., 2002;[140] 3 = Tailor et al., 2003;[171] 4 = Ekerhovd et al., 2001;[90] 5 = Canis et al., 1997;[80] 6 = Wakahara et al., 2001;[187] 7 = Maggino et al., 1994;[135] 8 = Schelling et al., 2000;[161] 9 = Roman et al., 1997;[42] 10 = Brown et al., 1998;[77] 11 = Granberg et al., 1990;[101] 12 = Hermann et al., 1987;[108] 13 = Kurjak and Predanic, 1992;[125] 14 = Tingulstad et al., 1996;[180] 15 = Stein et al., 1995;[168] 16 = Torres et al., 2002;[181] 17 = Manjunath et al., 2001;[139] 18 = Ma et al., 2003;[134] 19 = Valentin et al., 2001;[185] 20 = Franchi et al., 1995;[95] 21 = Merce et al., 1998;[146] 22 = Davies et al., 1993;[87] 23 = Morgante et al., 1999;[147] 24 = Benjapibal et al., 2003;[74] 25 = Gadducci et al., 1988;[97] 26 = Buy et al., 1996;[79] 27 = Strigini et al., 1996;[169] 28 = Luxman et al., 1991;[133] 29 = Kurjak et al., 1994;[127] 30 = Huber et al., 2002;[112] 21 = Reles et al., 1997;[155] 32 = Mancuso et al., 2004;[138] 33 = Kurjak et al., 2000;[124] 34 = Alcazar et al., 2003;[69] 35 = Kurjak et al., 1992;[126] 36 = Komatsu et al., 1996;[122] 37 = Yamashita et al., 1995;[192] 38 = Sohaib et al., 2005[166] 39 = Cohen et al., 2001;[33] 40 = Medl et al., 1995;[144] 41 = Hata et al., 1992;[106] 42 = Schneider et al., 1993;[43] 43 = Weiner et al., 1992;[188] 44 = Jain = 1994[117] 45 = Buist et al., 1994[78] 46 = Alcazar et al., 2003[67] 47 = Lin et al., 1993[39] 48 = Jain et al., 1993;[118] = Bromley et al., 1994;[76] 50 = Zimmer et al., 2003[195]

Comparing the figures, studies using the Sassone criteria show greater variability in sensitivity compared to variability in specificity (Figure 6), while those using the DePriest criteria (Figure 7) show greater variability in specificity and a relatively narrow range of sensitivity. Figure 10, which depicts a variety of other studies, suggests trade-offs between sensitivity and specificity; different morphology methods for discriminating benign from malignant have different thresholds, resulting in the sensitivity/specificity trade-off.

Three articles compared different scoring systems within the same study population. Caruso et al.[83] examined 112 women with adnexal masses comparing Sassone, DePriest, and Valentin scores. All performed similarly, displaying a sensitivity and NPV of 1.00, a range of specificity of 0.61 to 0.75, and a range of PPV of 0.35 to 0.48. Alcazar et al.[69] also compared the performance of Sassone, DePriest, and Ferrazzi. There were no significant differences between these scoring systems when receiver operating characteristic (ROC) curves were compared. The area under the curve (AUC) was 0.89 for Sassone, 0.92 for DePriest, and 0.90 for Ferrazzi. Ferrazzi et al.[93] evaluated 261 masses collected in three different centers. They compared ROC curves for scores based on Sassone, Granberg, DePriest, and Lerner's criteria and compared it with a scoring system they developed. The AUC ranged from 0.72 to 0.75 for the previously established systems. Their new scoring system (Ferrazzi) performed better, with an AUC of 0.84 ($p < 0.0001$). However, subsequent comparisons have not reaffirmed its superior functioning. When the Ferrazzi scoring system was compared to both Sassone and DePriest,[69] its performance was almost identical.

In spite of different designs, all the scoring systems performed similarly when compared within the same study population. It has been suggested that the poor performance of scoring systems with regard to their PPV is due to the misclassification of dermoid tumors.[197] Dermoids share many of the features that are characterized as "malignant" in scoring systems. The Alcazar study proposes a scoring system that was developed in part to correct this. Although this scoring system does perform well in its initial application, it has not been independently verified. The authors conclude, "a completely reliable differentiation of malignant masses cannot be obtained by sonographic imaging alone."[69]

Stratification by menopausal status. Of the 69 articles identified that addressed the assessment of adnexal morphology by ultrasound, only 13 contained data that either directly reported test characteristics by menopausal status or contained enough information to enable the stratification of results. Six were studies in a 100 percent postmenopausal patient population. Seven were studies that allowed comparison by menopausal status within the study population. They are presented in Table 11. The only significant difference in test performance appears to be in regards to the PPV. With the exception of Roman et al.,[42] the PPV is slightly higher in postmenopausal women. This likely reflects the higher prevalence of ovarian malignancy after menopause. Aside from the PPV, the performance of ultrasound in the morphological assessment of adnexal masses does not appear to be significantly changed by menopausal status.

Table 11. Ultrasound morphology assessment comparing pre- and postmenopausal status

Study	Scoring System	Premenopausal				Postmenopausal			
		Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV
Finkler et al., 1988[56]	Finkler	0.50	0.96	0.50	0.77	0.78	0.92	0.94	0.75
Franchi et al., 1995[95]	Descriptive	0.73	0.86	0.44	0.95	0.89	0.75	0.82	0.83
Guerriero et al., 2002[105]	Descriptive	0.98	0.89	0.44	1.00	1.00	0.51	0.52	1.00
Reles et al., 1997[155]	Modified score	1.00	0.79	0.46	1.00	0.87	0.89	0.77	0.94
Roman et al., 1997[42]	Descriptive	0.93	0.92	0.66	0.99	0.81	0.62	0.54	0.86
Schelling et al., 2000[161]	Descriptive	0.91	0.84	0.29	0.99	1.00	0.73	0.62	1.00

Study	Scoring System	Premenopausal				Postmenopausal			
		Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV
Alcazar et al., 2003[69]	Sassone DePriest Ferrazzi Alcazar	1.00 1.00 1.00 1.00	0.88 0.80 0.84 0.96	0.50 0.38 0.43 0.75	1.00 1.00 1.00 1.00	0.61 1.00 0.82 1.00	0.88 0.82 0.82 0.94	0.81 0.82 0.79 0.93	0.73 1.00 0.85 1.00
Menon et al., 2000[145]	Descriptive	-	-	-	-	1.00	0.94	0.24	1.00
Schutter et al., 1994[62]	Finkler	-	-	-	-	0.88	0.64	0.65	0.88
Bromley et al., 1994[76]	Unique scoring	-	-	-	-	0.91	0.52	0.52	0.92
Schutter et al., 1998[63]	Finkler	-	-	-	-	0.86	0.70	0.65	0.89
Luxman et al., 1991[133]	Descriptive	-	-	-	-	0.93	0.55	0.45	0.95
Kuriak et al., 1992[126]	Unique scoring	-	-	-	-	0.48	0.98	0.93	0.78

Abbreviations: NPV = negative predictive value; PPV = positive predictive value; Sens = sensitivity; Spec = specificity

Ultrasound Doppler Studies

Color Doppler scanning allows the assessment of tumor vascularity. Malignant neoplasms have active blood vessel creation (angiogenesis) compared to normal or benign neoplasms due, in part, to their increased metabolic activity. Overall, malignancies display an increased vascularity with decreased peripheral blood flow resistance and increased blood flow velocity compared with benign tissue.[152,200] Doppler signal analysis can separate high-resistance and low-resistance vessels and has therefore been investigated as a separate test modality, as well as in combination with ultrasound morphological evaluation in the evaluation of adnexal masses.

The most common flow criteria are the resistance index (RI), the pulsatility index (PI), and the maximum systolic velocity. RI is defined as the difference between peak systolic and maximum enddiastolic flow velocity, divided by peak systolic flow velocity. Usually the lowest measured RI from a series of measurements is reported from different arteries. PI is defined as the difference between peak systolic and enddiastolic flow velocity, divided by the time-averaged flow velocity. The maximum systolic velocity is the maximum flow recorded in any visualized artery.

In order to make a measurement of either RI or PI or maximum systolic velocity, an artery must be identified on ultrasound. The inability to identify an artery in the mass means that the test cannot be performed. Therefore, not every individual included in the study population is captured with the assessment of these color Doppler modalities. Another limitation of these measurements is that the range observed in malignant masses overlaps with that observed in benign masses. For example, in Lin et al.,[132] discussed in more detail below, the RI for malignant masses ranged from 0.23 to 0.82. Although they did not report a range for the benign masses, there were eight benign tumors with a RI < 0.4. This overlap limits the effectiveness of any threshold and, perhaps, contributes to the different thresholds reported in the literature.

Reproducibility of tests. Timmerman et al.[196] (discussed above under ultrasound morphology) included Doppler measurements in its analysis of interobserver variability and experience. In short, operators with more experience (300 versus 15,000 scans) had greater

accuracy (92 percent versus 82 to 87 percent, p = 0.0001). Interassessor agreement was also greater between the most experienced operators (kappa = 0.852) compared with the less experienced operators (range 0.581 to 0.737). None of the articles evaluating color Doppler described operator experience, nor did any address interobserver variability specifically in regards to Doppler measurement.

Trials identified. Fifty-six articles were identified that described color Doppler analysis, comprising a description of 65 tests. Thirty-two articles evaluated RI, 20 PI, and six the maximum systolic velocity. These are the most common flow criteria measured in the literature and presumably in clinical practice as well. Other Doppler parameters were described in the literature sometimes in conjunction with either RI or PI or maximum systolic velocity but were not included in this table. The other articles included 10 that involved the visualization of flow within the mass,[70,71,104,105,119,137,160,161,168,182] two that involved counting the total number of arteries (either > 4[152] or > 3[199]), and one that measured the absence of a diastolic notch.[137]

Results. Table 12 details the test characteristics of RI, PI, and the maximum systolic velocity in the evaluation of an adnexal mass, again using pooled values from a random-effects model. For RI the range reported was from ≤ 0.8 to < 0.4, with < 0.4 being the most common. For PI the range was relatively narrower from < 1.5 to < 1.0 with the majority of studies using either ≤ 1.0 or < 1.0. The reported range was greatest in the assessment of maximum systolic velocity, where there were also the fewest studies from > 30 cm/second to > 10 cm/second. As the threshold for RI decreases from ≤ 0.8 to < 0.4, the sensitivity and NPV decrease, and the specificity and PPV increase. This is seen most clearly in studies that evaluated a series of RI cutpoints with the same study population.[132,176]

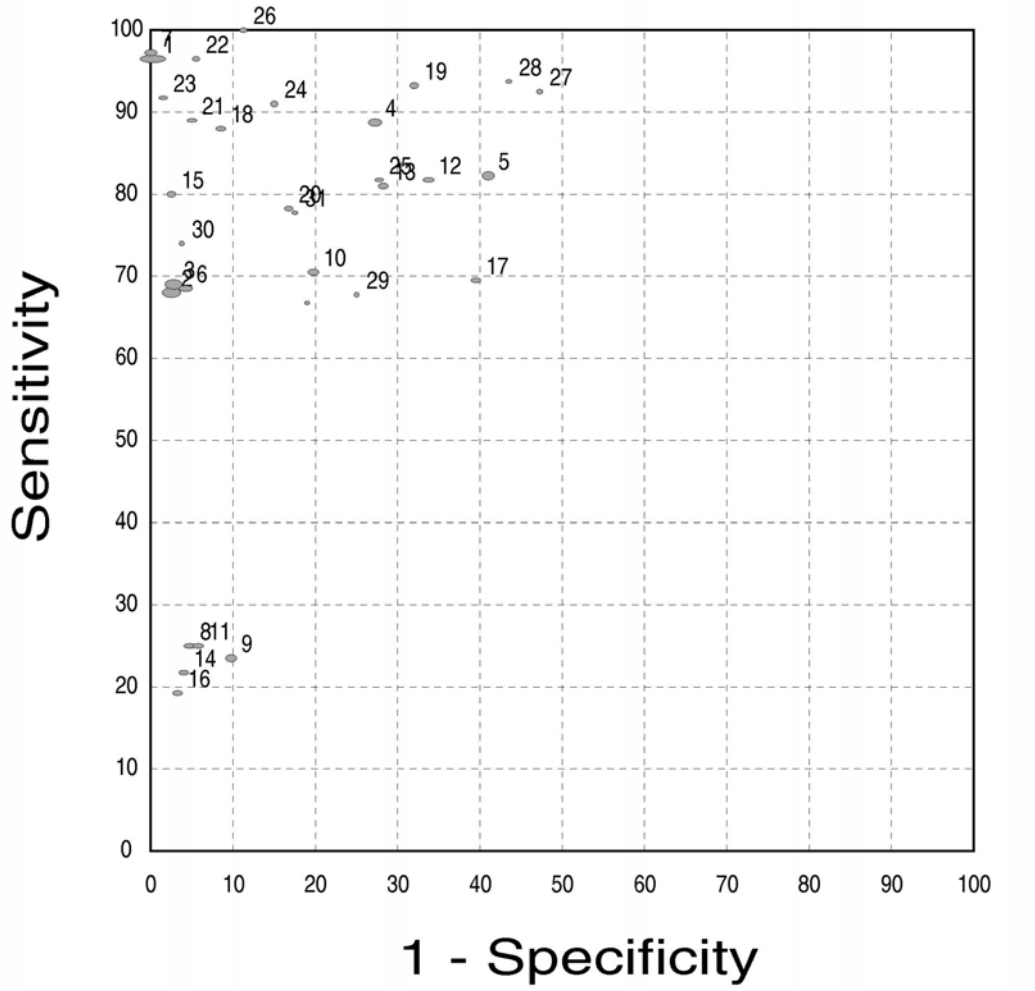
Lin et al.[132] evaluated 370 women with adnexal masses who were scheduled for surgery at a single institution. They reported outcomes based on RI cutpoints of 0.4, 0.5, and 0.6. For RI < 0.4, the sensitivity, specificity, PPV, and NPV were 0.69, 0.97, 0.89, and 0.91, respectively. For RI < 0.5, they were 0.79, 0.92, 0.77, and 0.93. And for < 0.6, they were 0.91, 0.86, 0.68, and 0.98. The authors conclude that the 0.4 cutpoint yields the highest concordance rate between Doppler prediction and histopathologic diagnosis. This conclusion, however, is based more on clinical impression, as ROC curve analysis was not performed.

The range of Doppler study performance is listed in Table 12 and shown in Figures 11-13. Overall there was great heterogeneity of performance results. The range of sensitivity was largest for RI. This range did not appear to be secondary to differences in RI thresholds; however, the < 0.4 threshold did appear to narrow specificity results. In spite of the large variation in thresholds described for maximum systolic velocity, the range of test characteristics was somewhat narrower than that for RI, probably because there were fewer studies identified that used this measurement. Again, there is a trade-off between sensitivity and specificity, although this appears greatest for maximum velocity.

Table 12. Sensitivity and specificity of Doppler studies

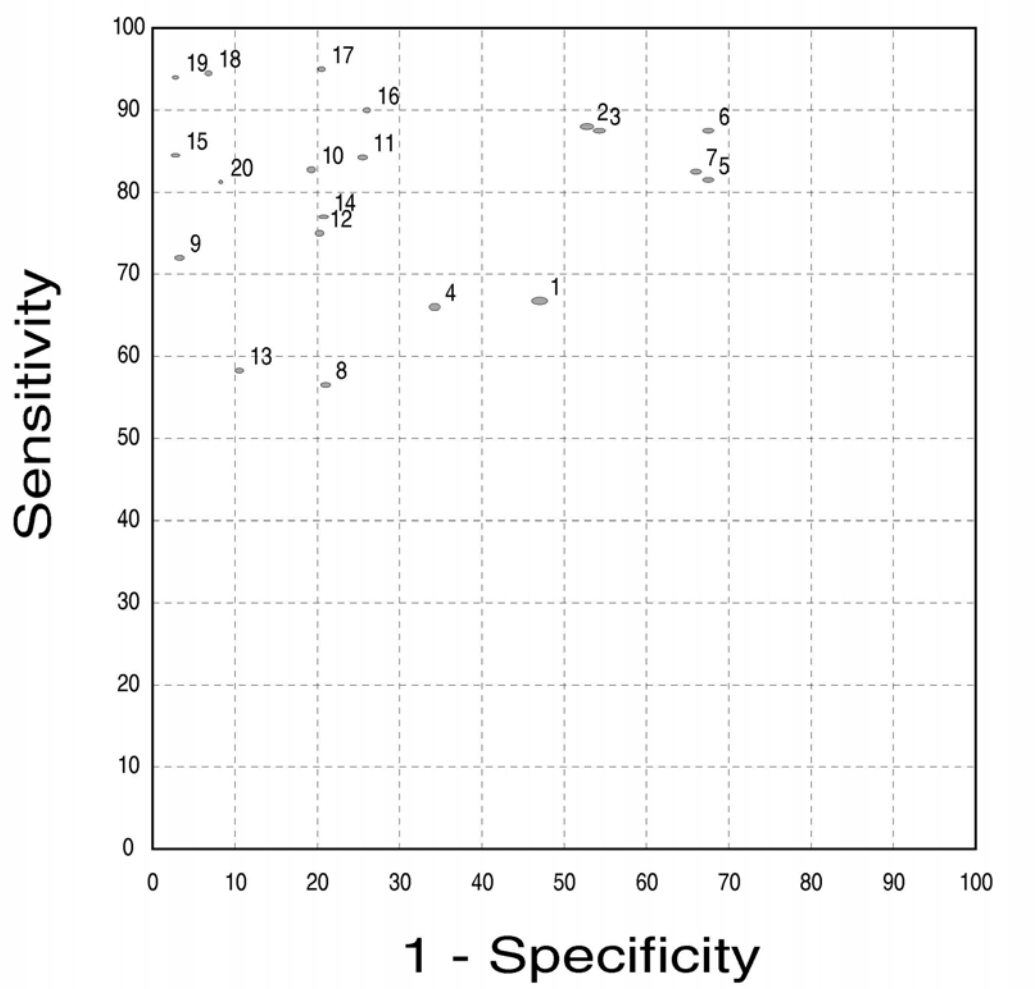
Doppler method	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Range of sensitivity in individual studies	Range of specificity in individual studies	References
Resistance index	0.76 (0.68 to 0.73)	0.89 (0.84 to 0.92)	0.19 to 1.00	0.53 to 1.00	43,68,70,75,76,79,81,86,88,95,106,107,117,124-126,128,130,132,141,146,152,168,172,175,176,179,184,190,193,199,201
Pulsatility index	0.79 (0.73 to 0.83)	0.74 (0.64 to 0.81)	0.57 to 0.95	0.32 to 0.97	73,79,81,94,103,109,115,120,154,155,158,163,168,169,179,182,184,188,199,201
Maximum systolic velocity	0.76 (0.61 to 0.86)	0.83 (0.66 to 0.93)	0.48 to 0.94	0.43 to 0.97	68,79,107,109,152,199

Figure 11. Performance of Doppler ultrasound resistance index



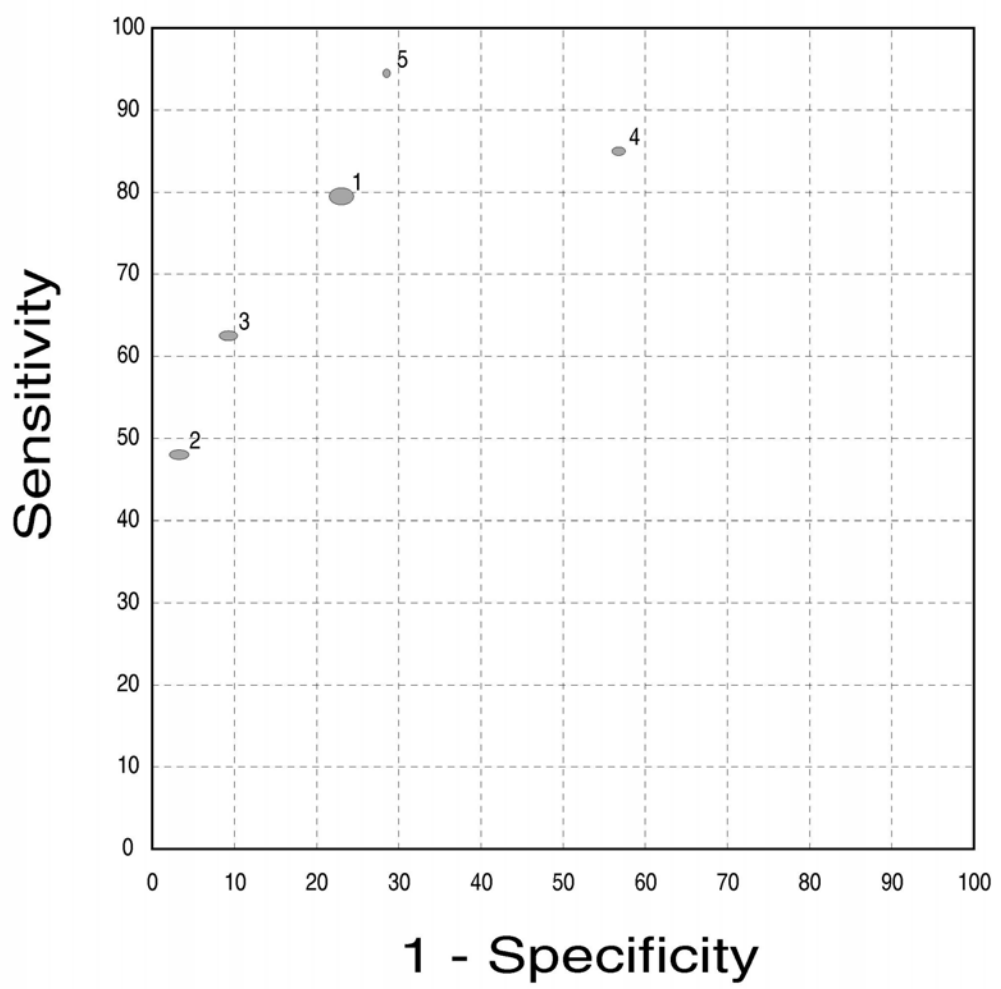
Key to Figure 11: 1 = Kurjak et al., 1991;[128] 2 = Wu et al., 1994;[190] 3 = Lin et al., 1993;[132] 4 = DePriest et al., 1994;[88] 5 = Prompeler et al., 1996;[152] 6 = Tepper et al., 1995;[176] 7 = Kurjak and Predanic, 1992;[125] 8 = Valentin, 2000;[184] 9 = Stein et al., 1995;[168] 10 = Anandakumar et al., 1996;[70] 11 = Valentin, 1996[201] 12 = Franchi et al., 1995;[95] 13 = Merce et al., 1998;[146] 14 = Carter et al., 1995;[81] 15 = Takac, 1998;[172] 16 = Buy et al., 1996;[79] 17 = Leeners et al., 1996;[130] 18 = Chou et al., 1994;[86] 19 = Hata et al., 1995;[107] 20 = Marret et al., 2004;[141] 21 = Kurjak et al., 2000;[124] 22 = Kurjak et al., 1992;[126] 23 = Timor-Tritsch et al., 1993;[179] 24 = Zanetta et al., 1994;[193] 25 = Tekay and Jouppila, 1992;[175] 26 = Alcazar et al., 1996;[199] 27 = Hata et al., 1992;[106] 28 = Schneider et al., 1993;[43] 29 = Berland et al., 2002;[75] 30 = Alcazar and Lopez-Garcia, 2001;[68] 31 = Jain, 1994;[117] 32 = Bromley et al., 1994[76]

Figure 12. Performance of Doppler ultrasound pulsatility index



Key to Figure 12: 1 = Rehn et al., 1996;[154] 2 = Guerriero et al., 1998;[103] 3 = Valentin, 2000;[184] 4 = Stein et al., 1995;[168] 5 = Itakure et al., 2003;[115] 6 = Valentin, 1999;[201] 7 = Valentin, 1997[182] 8 = Carter et al., 1995[81] 9 = Buy et al., 1996;[79] 10 = Benjapibal et al., 2002;[73] 11 = Kawai et al., 1994;[120] 12 = Strigini et al., 1996;[169] 13 = Hillaby et al., 2004;[109] 14 = Salem et al., 1994;[158] 15 = Timor-Tritsch et al., 1993;[179] 16 = Reles et al., 1997;[155] 17 = Alcazar et al., 1996;[199] 18 = Fleischer et al., 1992;[94] 19 = Weiner et al., 1992;[188] 20 = Sengoku et al., 1994[163]

Figure 13. Performance of Doppler ultrasound velocity indices



Key to Figure 13: 1 = Prompeler et al., 1996;[152] 2 = Buy et al., 1996;[79] 3 = Hillaby et al., 2004;[109] 4 = Alcazar et al., 1996;[199] 5 = Alcazar and Lopez-Garcia, 2001[68]

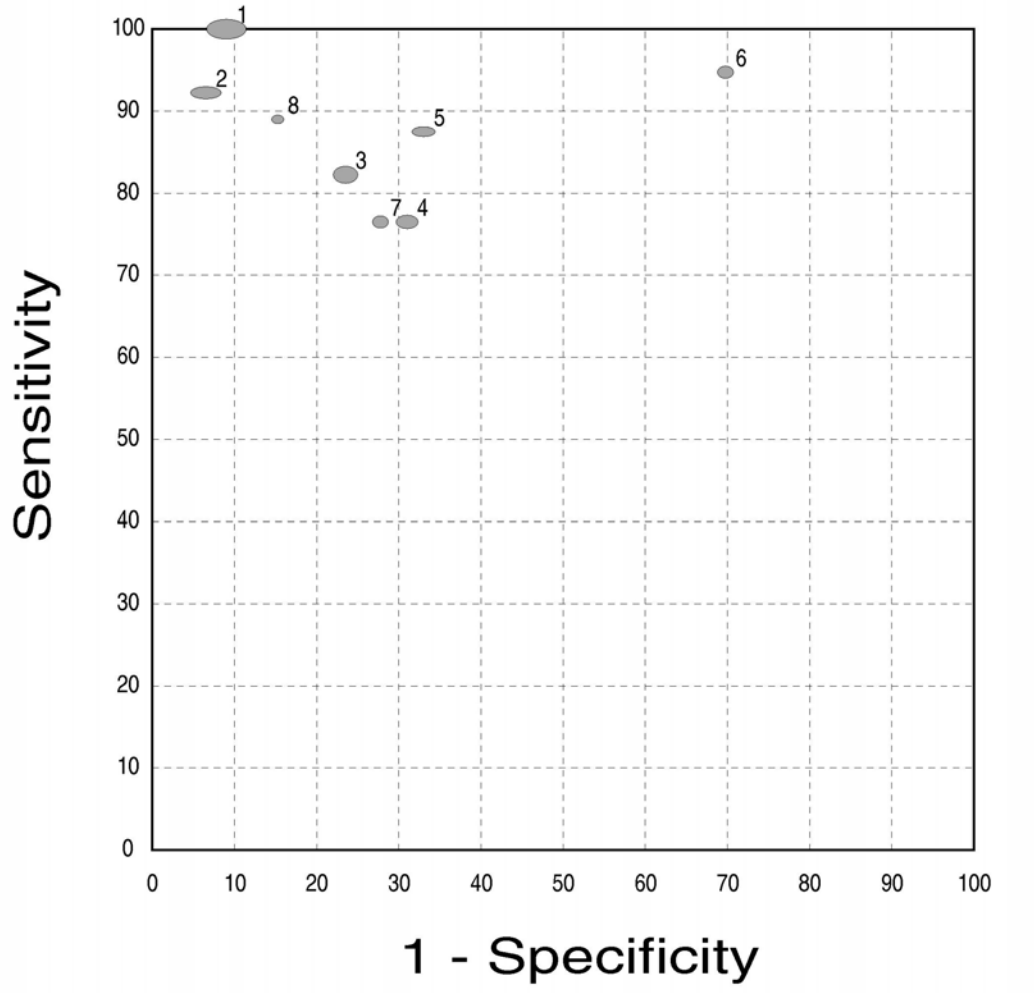
Table 13 compares the characteristics of Doppler studies that did not use measurement or calculation of Doppler waveforms. They relied instead on either the presence of vascularity within the mass (yes/no) or on a direct count of vessels seen. These tests seem to perform as well as the RI or PI in terms of sensitivity, although specificity varies quite widely (Figure 13). Valentin[182] measured both the PI (< 1.0) and the presence of color lakes visible on Doppler in the same study population. Of 151 patients, PI was measured in 135, indicating that for 16 individuals, no artery was visualized within the mass. The sensitivity reported for the PI was 0.83, specificity 0.34, PPV 0.20, and NPV 0.91. Simply documenting the presence or absence of visible color lakes on Doppler yielded a sensitivity of 0.88, a specificity of 0.67, a PPV of 0.33, and a NPV of 0.97. Not only did the direct visualization test perform better, but because its outcome was a simple binary outcome (present or absent), the results included the entire study population (n = 151). Prompeler et al.[152] measured RI, maximum systolic velocity, as well as the number of arteries visualized in the mass. Their data for the simple counting of arteries also performs as well if not better than the calculated tests such as RI or PI. In a random-effects

model, pooled sensitivity for the presence or absence of blood flow within a mass was 0.88 (95% CI, 0.80 to 0.92) and pooled specificity 0.78 (95% CI, 0.65 to 0.87)

Table 13. Study characteristics of simple Doppler visualization

Study (N)	Test	Sensitivity	Specificity
Prompeler et al., 1996[152] (212)	Total number of arteries > 4 (postmenopausal women only)	0.82	0.92
Valentin, 1997[182] (151)	Color lakes visible on Doppler	0.88	0.67
Maly et al., 1995[137] (102)	Demonstrable blood vessels	0.95	0.30
Schelling et al., 2000[161] (257)	Central vascularity on Doppler in solid component	0.93	0.94
Stein et al., 1995[168] (170 masses)	Internal flow within solid component or septation	0.77	0.69
Guerriero et al., 2002[105] (826 masses)	Arterial flow visualized in an echogenic structure or irregular solid portion	0.95	0.92
Anandakumar et al., 1996[70] (146)	"Continuously fluctuating" vessels with turbulent flow	0.77	0.68
Antonic and Rakar, 1995[71] (71)	Color flow present	0.89	0.47
Guerriero et al., 2005[104] (424)	Color flow present in "echogenic structure"	1.00	0.91
Juhasz et al., 1990[119] (147)	Color flow present in mass	0.96	0.84

Figure 14. Performance of Doppler ultrasound for intratumoral blood flow



Key to Figure 14: 1 = Guerriero et al., 2005;[104] 2 = Schelling et al., 2000;[161] 3 = Prompeler et al., 1996;[152] 4 = Stein et al., 1995;[168] 5 = Valentin, 1997;[182] 6 = Maly et al., 1995;[137] 7 = Anandakumar et al., 1996;[70] 8 = Antonic and Rakar, 1995[71]

Stratification by menopausal status. Out of a total of 56 studies identified that evaluated color Doppler, only 11 contained data that either directly reported test characteristics by menopausal status or contained enough information to enable the stratification of results. Two of these studies were in a 100 percent postmenopausal population, and nine enabled comparison by menopausal status within the same study population (Table 14). When comparing test performance within the same study population stratified by menopausal status, the PPV of the test is significantly increased in the postmenopausal group. In Salem et al.,[158] the PPV increased only from 0.20 in the premenopausal group to only 0.47 in the peri- and postmenopausal group. This may be a reflection of how they defined peri- and postmenopause (which was not clearly stated by the authors). After stratifying the reported results by age (> 45), the PPV is 0.73. This increase in PPV among postmenopausal women appears to be greater in the context of Doppler studies than that observed with ultrasound morphology. This finding differs from the one meta-analysis on the subject. Kinkel et al.[197] did a systematic review of both ultrasound morphology and Doppler in the detection of malignant masses. Although they noted a difference in outcomes

dependent on menopausal status, this difference did not reach statistical significance. Interestingly, there was a difference in terms of Doppler test performance by year of publication with better results demonstrated by earlier studies (p = 0.005), a result that was independent of sample size.

Table 14. Doppler studies stratified by menopausal status

Study (N)	Test	Premenopausal				Postmenopausal			
		Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV
Franchi et al., 1995[95] (129)	RI < 0.65	0.82	0.72	0.31	0.96	0.86	0.75	0.82	0.83
Guerriero et al., 2002[105] (826 masses)	Arterial flow visualized in echogenic structure or irregular solid portion	0.94	0.96	0.67	1.00	0.96	0.77	0.69	0.97
Reles et al., 1997[155] (98)	PI ≤ 1.1	0.80	0.67	0.36	0.93	0.93	0.83	0.76	0.91
Schelling et al., 2000[161] (257)	Presence of central vascularization on Doppler	0.91	0.94	0.53	0.99	0.93	0.92	0.84	0.97
Prompeler et al., 1996[152] (212)	Total number of arteries > 4 RI > 0.5 Maximum systolic velocity > 30cm/s	0.85 0.84 0.92	0.71 0.47 0.65	0.36 0.23 0.33	0.96 0.94 0.98	0.82 0.82 0.76	0.82 0.69 0.88	0.76 0.66 0.82	0.86 0.84 0.84
Strigini et al., 1996[169] (109)	PI < 1	0.83	0.73	0.21	0.98	0.85	0.81	0.73	0.90
Salem et al., 1994[158] (109 masses)	PI < 1	1.00	0.84	0.20	1.00	0.73	0.71	0.47	0.88
Szpurek et al., 2004[170] (464)	Doppler subjective index ≥ 4	0.82	0.93	0.79	0.94	0.92	1.00	1.00	0.82
Kurjak et al., 1992[126] (83)	RI < 0.41 randomly separate vessels	- -	- -	- -	- -	0.96 0.90	0.95 0.98	0.90 0.96	0.98 0.95
Bromley et al., 1994[76] (33)	RI < 0.6	-	-	-	-	0.66	0.81	0.67	0.81
Antonic and Rakar, 1995[71] (71)	Presence of color flow	1.00	0.36	0.11	1.00	0.87	0.79	0.81	0.85
Guerriero et al., 1998[103] (192 masses)	PI ≤ 1	0.86	0.46	0.08	0.98	0.88	0.52	0.66	0.81

Abbreviations: NPV = negative predictive value; PI = pulsatility index; PPV = positive predictive value; RI = resistance index; Sens = sensitivity; Spec = specificity

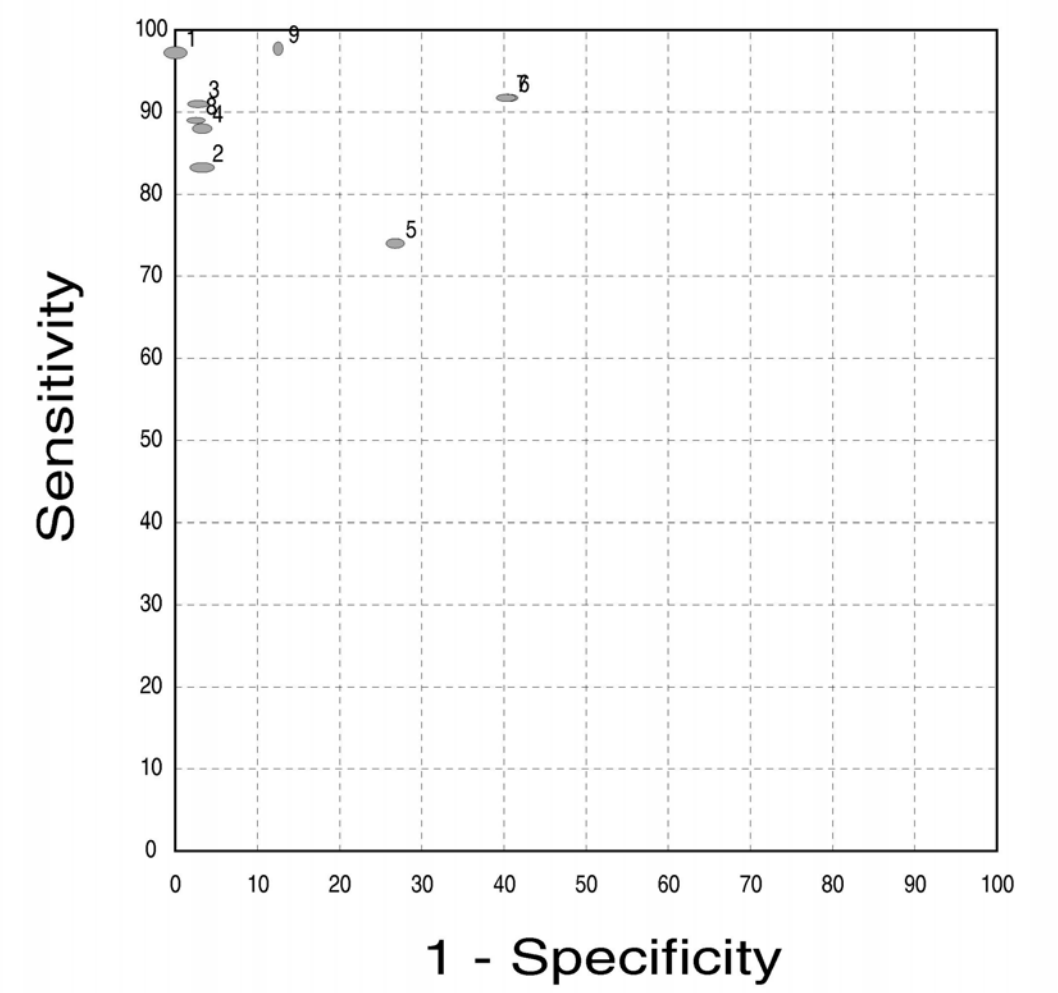
Combined Ultrasound Morphology and Doppler

A limiting feature of ultrasound morphologic assessments has been felt to be the high rate of false positive test results.[196] Color Doppler, in contrast, has displayed a slightly higher PPV, especially in the earlier studies.[197] There have, therefore, been attempts to combine ultrasound morphology and Doppler studies in a single test.

Trials identified. Of all the articles that investigated the use of either ultrasound morphology or color Doppler in the evaluation of an adnexal mass, nine articles containing a total of 13 tests described a combination ultrasound morphology and Doppler modality.[65,79,91,100,123-125,130,201]

Results. There is a large range in the reported study performance (sensitivity ranges from 0.71 to 0.98, specificity from 0.6 to 1.0. The relevant studies are shown in Figure 15; all but two had both sensitivity and specificity above 0.80. Pooled sensitivity in a random-effects model was 0.89 (95% CI, 0.81 to 0.93) and pooled specificity 0.91 (0.80 to 0.96). Both of these values were higher than the pooled values for any morphology or Doppler method alone.

Figure 15. Performance of combined ultrasound morphology and color Doppler



Key to Figure 15: 1 =Kurjak and Predanic, 1992;[125] 2 =Valentin, 1999;[201] 3 =Kurjak and Kupesic, 1999;[123] 4 =Buy et al., 1996;[79] 5 =Leeners et al., 1996;[130] 6 =Grab et al., 2000;[100] 7 =Fenchel et al., 2002;[91] 8 =Kurjak et al., 2000;[124] 9 = Alcazar and Castillo, 2005[65]

Stratification by menopausal status. There were two studies that analyzed combined ultrasound morphology and Doppler in 100 percent post menopausal patient populations. Kurjak et al.[126] reported a combined sensitivity, specificity, PPV, and NPV of 0.90, 0.94, 0.90, and 0.94, respectively. Their combined test consisted of RI < 0.41 and an ultrasound morphology scoring

system unique to them. Veunto et al.[186] in a population-based screening study reported a sensitivity, specificity, PPV, and NPV of 1.00, 0.83, 0.006, and 1.00, respectively. Given that these two studies are of greatly different design, it is hard to compare them directly. Comparing Kurjak et al. to the range of combined ultrasound and Dopper studies, it appears that in the postmenopausal group, the test has a better performance. However, this test performance may reflect patient selection criteria for the study that was not clearly explained. Combination modalities as a screening tool for ovarian cancer had a high false positive rate (as seen in the PPV of 0.006[186]).

3D Versus 2D Ultrasound

We identified five studies that analyzed 3D ultrasound. Four are listed in Table 15. The fifth, by Cohen et al.,[33] was not included because it compared 2D ultrasound with 2D plus some component of a 3D exam (possibly 3D Doppler) that was not clearly stated in the article. Overall, 3D ultrasound appears superior to 2D especially in regards to sensitivity and PPV performance. We were unable to stratify these results by menopausal status. Test reliability and variability were not addressed specifically in terms of 3D ultrasound.

Table 15. 3D versus 2D ultrasound

Study (number of persons)	Test	Sensitivity	Specificity	PPV	NPV
Alcazar et al., 2003[67] (41 masses)	2D 3D Presence of one of the following fulfilled criteria for mass: > 3 mm wall, > 3 mm septum, > 3 mm papillary projections, solid areas or echogenicity	0.90 1.00	0.61 0.78	0.68 0.81	0.88 1.00
Kurjak and Kupesic, 1999[123] (120)	2D 3D Both used a unique scoring system that included Doppler measurements	0.91 1.00	0.97 0.99	0.77 0.92	0.99 1.00
Kurjak et al., 2000[124] (90)	2D morphology 2D Doppler 2D combined 3D morphology 3D Doppler 3D combined Both used a unique scoring system for morphological assessment. Doppler for 2D was RI ≤ 0.42, for 3D it was “complex” “chaotic” vessel arrangement	0.67 0.89 0.89 0.78 0.89 1.00	0.94 0.95 0.98 0.98 0.98 0.99	0.55 0.67 0.80 0.78 0.80 0.90	0.96 0.99 0.99 0.98 0.99 1.00
Alcazar and Castillo, 2005[65] (69 masses)	2D 3D Presence of at least one of the following fulfilled criteria for “complex mass”: >3mm wall, > 3 mm papillary projection, solid areas or purely solid echogenicity Doppler flow in mass also used in test but unclear how	0.98 0.98	0.88 0.79	0.94 0.90	0.96 0.95

Abbreviations: 2D = two-dimensional; 3D = three-dimensional; NPV = negative predictive value; PPV = positive predictive value

Other Imaging Modalities

Although ultrasound remains the most common imaging modality in the evaluation and diagnosis of adnexal masses, newer technologies such as MRI, CT, and positron emission tomography (PET) have been studied as well. These modalities may not be as readily available to the clinician as ultrasound, and there is less literature devoted to them than to ultrasound; however, they are included in this review because of growing interest both clinical and research in their use. Further, despite refinements in ultrasound morphology scoring systems or Doppler measurements, the overall performance of ultrasound in the evaluation of the adnexal mass may be relatively fixed by the technology itself. Therefore it is necessary to investigate other imaging modalities and see how they compare with ultrasound.

Reproducibility of tests. Unlike ultrasound, MRI, CT, and PET images are not operator dependent in terms of obtaining the images. There is, however, the potential for interobserver variability in their analysis. There are no standardized morphological scoring systems for any imaging modality other than ultrasound. We identified two articles that directly addressed the issue of test reproducibility for either MRI and/or CT in the evaluation of adnexal masses. Buist et al.,[78] however, reported a series of 64 women who were evaluated by both MRI and CT and reviewed by two different radiologists. They reported a kappa value for the interobserver reliability for distinguishing between benign and malignant disease of 0.28 for CT and 0.41 for MRI. Yamashita et al.[192] also calculated kappa values for interobserver variability among five radiologists. They showed far greater agreement: for precontrast MRI, kappa = 0.71 (± 0.02); for contrast-enhanced MRI, kappa = 0.73 (± 0.02).

Trials identified. We identified 17 articles comprising 22 tests. There were 15 articles for MRI, three for CT, and three for PET and one that used a combined CT/MRI test. There were two articles that investigated nuclear medicine technologies in the evaluation of adnexal masses. These, however, were not included in the review given the experimental nature of such tests at this time. The PET studies were all performed also using tracer 18-Fluorodeoxyglucose (FDG) with the test measuring uptake of FDG in the lesion.

Results. The results of MRI, CT, and PET modalities are summarized in Table 16. All of the articles describing CT and PET and most of the articles describing MRI either used descriptive criteria for differentiating malignant from benign appearing lesions or did not report the criteria used. Only two articles for MRI used a scoring system, slightly different from each other, which increases the difficulty in comparing studies. To date, there are no standardized scoring systems for any imaging modality other than ultrasound.

The range of test performance of MRI, CT, and PET are shown in Table 16. Table 17 includes, for comparison, the test performance for ultrasound morphology, color Doppler (all the modalities), and ultrasound morphology and Doppler combined. Tian et al.[177] was excluded from this table because there was no description how CT and MRI were combined for a single test result (in series versus in parallel). Overall the sensitivity for MRI, CT, and PET are similar to that of combined ultrasound morphology and Doppler and less heterogenous than either modality separate. The specificity, however, is equivalent to either test separate and wider than the tests combined, with the exception of FDG-PET. However, the comparatively narrow range of both CT and PET results could be secondary to the relatively few studies that use these modalities. There is a large range of results for PET PPVs and a small range for CT, again possibly reflecting the paucity of studies. The range of NPVs for MRI is comparable to that for combined ultrasound morphology and Doppler and better than either CT or PET. Overall MRI

appears similar in performance to combined ultrasound. More research is needed to accurately assess the performance range of CT and PET.

Table 16. Sensitivity and specificity of other imaging modalities

Imaging modality	Pooled sensitivity (95% CI)	Pooled specificity (95% CI)	Range of sensitivity in individual studies	Range of specificity in individual studies	References
MRI	0.91 (0.86 to 0.94)	0.87 (0.83 to 0.90)	0.67 to 1.00	0.77 to 1.00	44,78,91,100,106,111,112,118,121,122,129,144,156,166,192
CT	0.90 (0.83 to 0.94)	0.75 (0.36 to 0.94)	0.86 to 0.96	0.35 to 0.89	39,78,129
FDG-PET	0.67 (0.52 to 0.79)	0.79 (0.70 to 0.85)	0.58 to 0.78	0.76 to 1.00	91,100,121

Abbreviations: CI = confidence interval; CT = computed tomography; FDG = 18-Fluorodeoxyglucose; MRI = magnetic resonance imaging; PET = positron emission tomography

Another way to compare imaging modalities is by looking at studies that compare imaging modalities within the same study population. These are listed in Table 17. There may be a small benefit in performance of MRI over ultrasound, especially in terms of PPV. There is no evidence to support the superiority of any single modality, although FDG-PET appears inferior to the rest.

Table 17. Comparison of MRI, CT, FDG-PET, and ultrasound

Study (N)	Test	Sensitivity	Specificity	PPV	NPV
Medl et al., 1995[144] (73)	Ultrasound morphology (descriptive) MRI descriptive	0.81 0.97	0.73 0.83	0.79 0.88	0.76 0.96
Yamashita et al., 1995[192] (72 women 80 masses)	Ultrasound morphology (unique score) MRI precontrast MRI contrast enhanced	0.89 0.78 0.91	0.84 0.93 0.93	0.63 0.79 0.81	0.96 0.93 0.97
Fenchel et al., 2002[91] (99)	Ultrasound combined morphology and Doppler MRI FDG-PET	0.92 0.83 0.58	0.60 0.83 0.76	0.24 0.40 0.25	0.98 0.97 0.93
Jain et al., 1993[118] (32)	Ultrasound morphology (descriptive) MRI	1.00 0.67	0.60 1.00	0.18 1.00	1.00 0.97
Kawahara et al., 2004[121] (38)	MRI descriptive FDG-PET	0.91 0.78	0.87 1.00	0.91 1.00	0.87 0.75
Komatsu et al., 1996[122] (82)	Ultrasound morphology (unique score) MRI descriptive (n = 59)	1.00 0.91	0.46 0.88	0.57 0.91	1.00 0.88
Lin et al., 1993[39] (80)	Ultrasound morphology (descriptive) CT descriptive	0.83 0.86	0.50 0.36	0.58 0.74	0.79 0.56
Buist et al., 1994[78] (64)	CT reviewer a CT reviewer b MRI reviewer a MRI reviewer b Ultrasound morphology (NR)	0.96 0.89 0.96 0.96 0.89	0.44 0.83 0.33 0.94 0.44	0.72 0.89 0.68 0.96 0.71	0.89 0.83 0.86 0.94 0.73
Grab et al., 2000[100]	Ultrasound combination morphology and Doppler	0.92 0.83	0.60 0.84	0.23 0.42	0.98 0.97

Study (N)	Test	Sensitivity	Specificity	PPV	NPV
(101)	MRI descriptive FDG-PET	0.58	0.80	0.28	0.93
Hata et al., 1992[106] (63)	Ultrasound (NR) MRI score	0.85 0.67	0.69 0.97	0.68 0.95	0.86 0.80
Huber et al., 2002[112] (93)	Ultrasound morphology (descriptive) MRI descriptive	0.85 0.89	0.73 0.86	0.87 0.93	0.71 0.79
Reuter et al., 1998[156] (65)	Ultrasound morphology (descriptive) MRI descriptive	1.00 1.00	0.66 0.78	0.40 0.50	1.00 1.00
Sohaib et al., 2005[166] (72)	Ultrasound morphology (descriptive) MRI descriptive	1.00 0.97	0.40 0.84	0.53 0.80	1.00 0.97

Abbreviations: CT = computed tomography; FDG = 18-Fluorodeoxyglucose; MRI = magnetic resonance imaging; NR = not reported; PET = positron emission tomography

Only two studies compared pre- and postcontrast enhancement with MRI.[111,192] Contrast enhancement improved evaluative performance in both studies, particularly sensitivity. In Hricak et al. the sensitivity increased from 0.87 to 0.95, specificity from 0.75 to 0.79, PPV from 0.78 to 0.83, and NPV 0.84 to 0.94.[111] These results are similar to those of Yamashita et al.[192] in Table 17.

Stratification by menopausal status. None of the studies describing MRI, CT, or PET reported results either by menopausal status or in data that would allow menopausal status to be stratified.

Serum Markers: CA-125

The concept of using tumor markers as either screening or diagnostic tests for ovarian cancer is dependent upon identifying an abnormal level of a particular marker in serum, reflecting a systemic effect of disease in the ovary. The most extensively investigated ovarian cancer associated antigen is CA-125. This antigen is recognized by a murine monoclonal antibody produced using an ovarian cancer cell line as an immunogen. Elevated levels are detected in approximately 80 percent of ovarian carcinomas at the time of diagnosis;[136,167] however, elevated serum levels have also been reported in a variety of benign conditions, potentially affecting specificity. In addition, CA-125 is not as commonly elevated in non-epithelial ovarian cancers. Because these stromal and germ cell tumors are proportionately more common in pre-menopausal women, the sensitivity of CA-125 may it is not as sensitive in premenopausal women.[3]

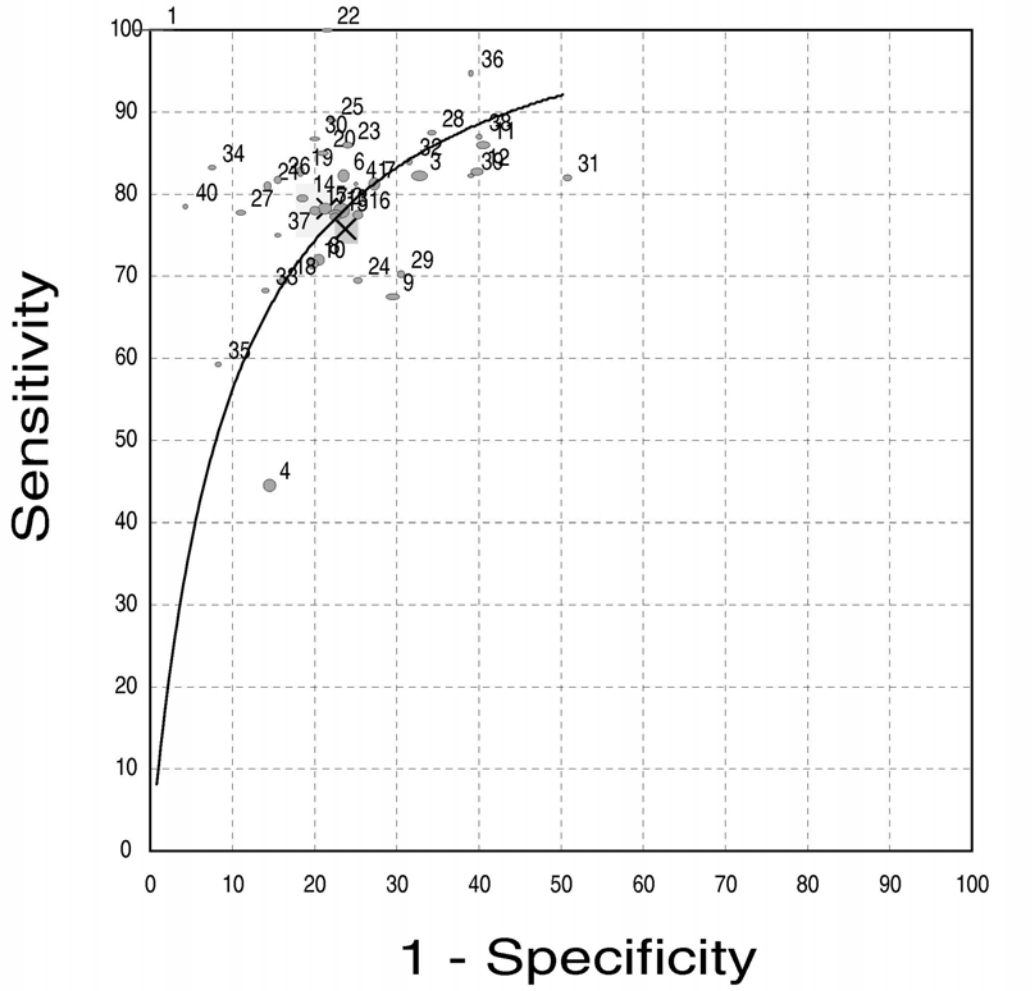
Reproducibility of tests. Only one study included specific information regarding the inter- and intra-assay coefficients of variation.[66] They were < 7.5 percent and < 5.3 percent, respectively. The sensitivity of the assay in this study was < 5 U/ml.

Trials identified. We identified 66 studies that investigated the use of CA-125 as a serum marker in the evaluation of an adnexal mass. One study was a population-based screening study that employed CA-125 as part of the screening triage.[51] Forty-six studies in total used 35 U/ml as a threshold – in 37 it was the only threshold used, whereas in five, both 35 U/ml and another threshold were reported for the same patient population. There were 24 studies that reported a threshold other than 35 U/ml ranging from >20 U/ml to >100 U/ml. In addition to the five

studies that reported 35 U/ml and an additional level, there were four other studies that reported two threshold levels within the same study population. All but one of the studies were case series. Although there were a few studies that compared CA-125 results from operative cases with normal controls, only the data from the operative series were included in the 2-by-2 tables. The clinical presentation of the cases was rarely described. Some of the series were drawn from oncology clinics

Results. At the most commonly used threshold of 35 U/mL, the pooled sensitivity of CA-125 for discriminating benign from malignant lesions was 0.78 (95% CI, 0.75 to 0.81), and the pooled specificity 0.78 (95% CI, 0.71 to 0.82). Individual study sensitivities ranged from 0.45 to 1.0, and specificities from 0.46 to 0.99; see Figure 16, where the trade-off between sensitivity and specificity resulting from different thresholds is clearly seen. Not including the one screening study in this series,[51] the studies ranged in size from 52 to 429 individuals. Unlike ultrasound morphology assessments, the range of CA-125 performance is not influenced by the heterogeneity of evaluative modalities. However, the results of performance have, overall, a similarly broad range. This most likely reflects heterogeneity of study populations. As very few studies actually reported how patients were diagnosed with masses, it is impossible to accurately stratify these results by patient characteristics. As with ultrasound measurements (both morphology and Doppler), the narrowest range of CA-125 test performance was with NPV, making this, perhaps, the most reliable part of the test itself.

Figure 16. Performance of CA-125



Key to Figure 16: 1 = Adonakis et al., 1996;[51] 2 = Woolas et al., 1995;[189] 3 = Gadducci et al., 1992;[98] 4 = Wakahara et al., 2001;[187] 5 = Maggino et al., 1994;[135] 6 = Dowd et al., 1993;[55] 7 = Schutter et al., 2002;[162] 8 = Patsner and Mann, 1988;[151] 9 = Roman et al., 1997;[42] 10 = Schutter et al., 1994;[62] 11 = Gadducci et al., 1991;[99] 12 = Chen et al., 1988;[85] 13 = Vasilev et al., 1988;[50] 14 = Timmerman et al., 1999;[178] 15 = Hogdall et al., 2000;[110] 16 = Malkasian et al., 1988;[136] 17 = Torres et al., 2002;[181] 18 = Schutter et al., 1998;[63] 19 = Manjunath et al., 2001;[139] 20 = Troiano et al., 1997;[47] 21 = Chalas et al., 1992;[31] 22 = Mancuso et al., 2004;[138] 23 = Gadducci et al., 1988;[97] 24 = Finkler et al., 1988;[56] 25 = Tay and Chua, 1994;[174] 26 = Soper et al., 1990;[167] 27 = Smikle et al., 1995;[46] 28 = Hurteau et al., 1995;[113] 29 = Asif et al., 2004;[72] 30 = Einhorn et al., 1986;[89] 31 = Hillaby et al., 2004;[109] 32 = Alcazar et al., 1999;[66] 33 = Balbi et al., 2001;[53] 34 = Antoni and Rakar, 1995;[71] 35 = Hata et al., 1992;[106] 36 = O'Connell et al., 1987;[148] 37 = Schneider et al., 1993;[43] 38 = Weiner et al., 1992;[188] 39 = Tian et al., 2000;[177] 40 = Berlanda et al., 2002;[75] 41 = Sengoku et al., 1994[163]

The only screening study identified for CA-125 in our literature search[51] included 2000 women. The sensitivity in this study was 1.00, specificity 0.99, PPV 0.17, and NPV 1.00. Few of the other studies achieved this degree of sensitivity, specificity, or NPV, although overall the PPV was higher. In the presence of an adnexal mass, the false negative rate increases compared with a screened population reflecting the fact that benign gynecologic disease can cause elevation of CA-125.

The most common threshold other than 35 U/ml was 65 U/ml. Most of the studies using 65 U/ml as a threshold were from Asia. The probable heterogeneity of study populations makes

comparisons between these levels limited. Looking at the studies that reported results for different levels of CA-125 for within the same study population,[87,98,134,136,147,148,162,167,180] in the higher threshold measurement, the specificity and PPV are higher, the sensitivity is lower, and the NPV is only slightly lower.

Stratification by menopausal status. Of the 59 studies we identified that examined CA-125, only nine contained data that either directly reported test characteristic by menopausal status or contained enough information to enable the stratification of results. One study was conducted exclusively in a postmenopausal population.[63] The studies are listed in Table 18.

The incidence of ovarian cancer is higher in postmenopausal women relative to benign gynecologic conditions, which also increase CA-125 levels. This should translate into a greater accuracy of CA-125 test performance in this population. Indeed, all test parameters except NPV are both higher and the range narrower in postmenopausal women. The lowest PPV was 0.73, with the remaining above 0.85, which is significantly higher than the range of PPV observed in studies that did not stratify their results by menopausal status. The NPV is lower in the postmenopausal population, despite the higher sensitivity, because of a greater prevalence of cancer in this population. CA-125 is consistently more helpful in discriminating benign from malignant lesions in postmenopausal women compared with premenopausal women.

Table 18. CA-125 results stratified by menopausal status

Study	Threshold	Premenopausal				Postmenopausal			
		Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV
Malkasian et al., 1988[136]	> 100 > 35	0.60 0.60	0.95 0.73	0.67 0.29	0.93 0.91	0.77 0.81	0.97 0.91	0.98 0.94	0.72 0.74
Gadducci et al., 1996[96]	> 65	0.67	0.91	0.67	0.91	0.80	1.00	1.00	0.69
Gadducci et al., 1992[98]	> 64	0.50	0.26	0.05	0.86	0.81	0.86	0.88	0.78
Franchi et al., 1995[95]	> 39	0.73	0.64	0.24	0.94	0.77	0.85	0.87	0.74
Patsner and Mann, 1988[151]	> 35	0.63	0.78	0.66	0.76	0.77	0.81	0.85	0.72
Dowd et al., 1993[55]	> 35	0.74	0.73	0.60	0.84	0.86	0.82	0.90	0.76
Finkler et al., 1988[56]	> 35	0.50	0.69	0.35	0.81	0.84	0.92	0.94	0.80
Schutter et al., 1998[63]	> 35	--	--	--	--	0.69	0.84	0.73	0.81
Antonic and Rakar, 1995[71]	> 35	0.67	0.92	0.40	0.97	0.87	0.93	0.93	0.87

Abbreviations: CA-125 = cancer antigen 125; NPV = negative predictive value; PPV = positive predictive value; Sens = sensitivity; Spec = specificity

Other Serum Markers

The fact that CA-125 is < 35 U/ml in 20 percent of women with early stage ovarian cancer, has motivated research into other serum based tests. We identified 13 articles that described a total of 17 different sera studies in women with an adnexal mass. Some studies investigated the performance of other tumor-associated antigens such as tumor-associated glycoprotein 72 (TAG-72) or CA-19-9. Although most of the tumor-associated antigens achieved specificities of

approximately 0.82 to 0.92, the sensitivity, PPV, and NPV were overall lower than those reported for CA-125. Two studies investigated carcinoembryonic antigen (CEA),[114,157] and although they employed slightly different thresholds, the sensitivity reported in both (0.16 and 0.22) are so poor as to lead both authors to conclude that assessment of CEA in the evaluation of an adnexal mass is not helpful. Roman et al.[42] investigated whether the addition of human chorionic gonadotropin (hCG), alpha-fetoprotein (AFP), and lactate dehydrogenase (LDH) to CA-125 improved the test performance. In their series the sensitivity of CA-125 alone was 0.67, the specificity was 0.71, PPV 0.35, and NPV 0.90. The addition of the other three tests did not change the test results very much. The combined test (defined as any of the markers positive) sensitivity was 0.72, its specificity was 0.70, PPV 0.36, and NPV 0.94. AFP, hCG, and LDH do not appear to improve the diagnostic performance of CA-125.

Gadducci et al. investigated the role of D-Dimer in a series of 121 women with adnexal masses.[96] The sensitivity for D-Dimer alone was 0.91, the specificity was 0.83, the PPV 0.82, and the NPV 0.92 – making D-Dimer one of the best performing tests identified in our review. Stratifying by menopausal status showed a greater performance in premenopausal women where the sensitivity, specificity, PPV and NPV were 1.00, 0.91, 0.75, and 1.00 respectively (n = 57). For postmenopausal women they were 0.89, 0.65, 0.85, and 0.72, respectively. Chalas et al. investigated the role of elevated platelets in 241 women.[31] The specificity and PPV were similar to that reported for D-Dimer (0.84 and 0.83, respectively), but the sensitivity and NPV were significantly lower (0.56 and 0.59). These two studies are intriguing, but the results need to be established in future studies to better assess their possible contribution to the evaluation of adnexal masses.

Aside from D-Dimer, none of the studies contained information making stratification by menopausal status possible. In conclusion, none of the sera markers investigated in this review appears to perform better than CA-125, with the possible exception of D-Dimer in the premenopausal population.

Population-based Studies

Almost all of the studies identified were case series. There were, however, 13 population-based screening studies included in this review. They are listed in Table 19. Although all of the women included in these studies did not have a diagnosis of an adnexal mass at the time of enrollment, these studies are included here because they highlight some important issues about test performance. The strongest studies from a methodological perspective were those by Marchetti et al.,[140] Vuento et al.,[186] DePriest et al.,[34] Adonakis et al.,[51] and Tailor et al.[171] Marchetti, Vuento, Tailor and DePriest all used ultrasound as a screening modality. In all of these studies, the PPV was low, ranging between 0.006 to 0.07. Screening with CA-125 yielded a slightly higher PPV of 0.17.[51] Tailor et al.[171] offered followup screening within the same populations. In the first screening episode, which captured the total study population of 2,500 women, the test characteristics were similar to those reported in the other screening studies. The test characteristics improved, however, with subsequent screening. Women who had a negative screen were offered either a 12- or 6-month repeat ultrasound (depending on individual risk factors). Nine hundred and ninety-eight women received a second ultrasound screening. For this subset, the PPV improved to 0.21. For women screened greater than two times, the PPV was 0.25. However, not all women offered additional screening returned for the ultrasound. This potential bias was not discussed by the authors, and it is unclear how it may have influenced the

performance of repeat screening. The three studies by Kurjak et al. each had various biases that could have accounted for their markedly different reported test performances. One did not report followup on test negatives and therefore included no false negative in the series,[126] another study population was an undescribed subset of a larger still incomplete screening series,[127] and the last study did not describe inclusion criteria.[128] Van Nagell et al.[49] screened 14,469 women with ultrasound. They reported their results 12 months from the time of screening. However they note that four women were diagnosed with cancer greater than 12 months after screening. These women had all screened negative and were included in their analysis as true negatives. Reclassifying these individuals as false negatives changes the sensitivity from 0.81 to 0.68.

Table 19. Population-based screening studies

Study	N	Test	Sensitivity	Specificity	PPV	NPV
Marchetti et al., 2002[140]	4350	Ultrasound screening: criteria NR Operative cases only (n = 45) Assuming all negatives were truly negative (n = 4359)	1.00 1.00	0.37 0.96	0.07 0.01	1.00 1.00
Menon et al., 2000[145]	1027	Ultrasound Volume > 8.8 ml Abnormal morphology Complex morphology	 0.90 1.00 0.84	 0.94 0.94 0.97	 0.21 0.24 0.37	 1.00 1.00 0.98
Vuento et al., 1995[186]	1364	Combined ultrasound morphology and Doppler (PI < 1.0)	1.00	0.88	0.006	1.00
DePriest et al., 1993[36]	24/3220	Ultrasound morphology (DePriest) Operative cases only (n = 24)	1.00	0.71	0.33	1.00
Kurjak et al., 1992[126]	83/1000	RI < 0.41 Ultrasound morphology (unique score) Presence of random vessels Combined ultrasound and Doppler	0.96 0.48 0.90 0.90	0.95 0.98 0.98 0.94	0.90 0.93 0.96 0.90	0.98 0.78 0.95 0.94
Kurjak et al., 1994[127]	32/5013	Ultrasound “persistent mass” Ultrasound assuming all test negatives true negatives	1.00 1.00	0.97 0.99	0.80 0.80	1.00 1.00
Kurjak et al., 1991[128]	680/ 8620	RI < 0.4	0.96	0.99	0.98	1.00
DePriest et al., 1997[34]	90/6470	Ultrasound morphology (DePriest) (n = 90) Assuming all test negatives true negatives (n = 6470)	1.00 0.86	0.59 0.99	0.17 0.07	1.00 1.00
Adonakis et al., 1996[51]	2000/ 2000	CA-125 > 35 PE “palpable mass”	1.00 0.67	0.99 0.97	0.17 0.03	1.00 1.00
Andolf et al., 1990[52]	801	Combined ultrasound and BME (both positive for test to be positive) Ultrasound and BME criteria not well described	1.00	0.94	0.11	1.00
Jacobs et al., 1988[58]	1010	CA-125 > 30 U/ml BME Ultrasound (ovarian volume > 8.8ml) (n = 58 for ultrasound)	1.00 1.00 1.00	0.97 0.97 0.74	0.03 0.04 0.08	1.00 1.00 1.00
Tailor et al., 2003[171]	2500	Ultrasound morphology (descriptive)	0.86	0.97	0.07	1.00

Study	N	Test	Sensitivity	Specificity	PPV	NPV
		N = 2500 Ultrasound for second screening episode (n = 998) Ultrasound for >= third screening episode (n = 733)	1.00 1.00	0.99 0.99	0.21 0.25	1.00 1.00
van Nagell et al., 2000[49]	14469	Ultrasound (ovarian volume > 20 cm^3 for premenopausal women, > 10 cm^3 for postmenopausal women)	0.81	0.99	0.09	1.00

Abbreviations: BME = bimanual examination; CA-125 = cancer antigen 125; NR = not reported; PE = pelvic examination; PI = pulsatility index

Methodological Issues

In reviewing the literature on evaluation modalities, numerous methodological problems consistently reduced our ability to draw conclusions about the performance of various tests both individually and in comparison with each other. Some of these problems concerned study design, others related to statistical issues.

Patient population. With the exception of the 13 population-based screening studies, all of the articles were case series. Some were consecutive and others non-consecutive. Some were based on operative cases within a specific time frame at one or several institutions, whereas others were referral series, often located in oncology clinics. The path to diagnosis was almost never described, making it difficult to asses the generalizability of the results. Further, age was the only patient characteristic that was reliably documented. Other characteristics, such as family history, were almost never included. This has several implications. The overrepresentation of operative cases especially from academic facilities, likely overrepresents the prevalence of malignancy in the study populations when compared with the population of women with adnexal masses in general. It also exaggerates the performance of the evaluative modalities, especially in regards to sensitivity and PPV. Finally, it limits the generalizability of the evidence.

Definition of malignant. There was inconsistency between studies regarding whether the malignant classification included any malignancy or whether it included only ovarian malignancies. The inclusion of all malignancies would exaggerate the test's specificity and PPV at the expense of its sensitivity and NPV. From a practical standpoint, this difference may not be that problematic, as all malignancies are important. However, this classification bias increased the heterogeneity of test performance and limits generalizability. Finally, almost all of the articles that reported series containing tumors of low malignant potential (LMP) (also called borderline) classified these tumors as malignant. This changes the reported performance of the various evaluative modalities in these studies. There were three studies identified where stratification by LMP was possible. These are listed in Table 20. Classifying LMP tumors as malignant increases the specificity and PPV relative to classifying them as benign, while decreasing the specificity and NPV. Overall, PPV tended be somewhat low (even in populations with high prevalences of disease). The inclusion of LMP tumors into the malignant category inflated this measurement somewhat. Obviously, because of uncertainty about the natural history of LMP tumors, the most appropriate way of classifying them as part of diagnostic test

evaluation is also uncertain. Given this uncertainty, ideally investigators would report results using alternative methods of classifying LMP tumors.

Table 20. Effect of classification of LMP tumors as malignant or benign on diagnostic test characteristics

Study	Test	LMP classified as malignant				LMP classified as benign			
		Sens	Spec	PPV	NPV	Sens	Spec	PPV	NPV
Roman et al., 1998[157]	CEA	0.16	0.93	0.35	0.83	0.19	0.93	0.25	0.90
Wakahara et al., 2001[187]	Ultrasound morphology CA-125	0.82 0.45	0.82 0.86	0.65 0.74	0.92 0.63	0.86 0.77	0.78 0.61	0.54 0.37	0.95 0.90
Timmerman et al., 1999[178]	CA-125	0.80	0.82	0.63	0.91	0.77	0.79	0.56	0.91

Abbreviations: CA-125 = cancer antigen 125; CEA = carcinoembryonic antigen; LMP = low malignant potential (tumors); NPV = negative predictive value; PPV = positive predictive value; Sens = sensitivity; Spec = specificity

Variability in test criteria. Of the 69 articles that evaluated ultrasound morphology, only 31 used established scoring criteria; 38 used a novel method. This resulted in a great heterogeneity of tests for ultrasound morphology and contributed to the range in performance noted. Many of the studies employed purely descriptive analysis to arrive at a benign versus malignant diagnosis. This limits the reproducibility of those results. Many of the scoring systems and descriptive categories had never been independently verified, and the paucity of details regarding what constituted a positive test makes such verification impossible. In terms of ultrasound evaluation by color Doppler, there was also a range of reported thresholds. Some of the variability in test criteria reflects the limitations of ultrasound technology. However, such differences limited the comparability between studies.

Masses as numerator. While most studies examined persons as the unit of 2-by-2 analysis, there were many studies that analyzed their data by masses. Even though the number of persons in the study was usually reported, it was often impossible to reconfigure the 2-by-2 table to refer to persons not masses. This was especially true in the radiology literature. This influenced the comparability between studies.

Menopausal status. Most of the studies did describe the patient population in terms of age. We were able to calculate the proportion of menopausal patients in most studies. However, the results were rarely reported in a way that allowed stratification by menopausal status. Where stratification was possible, a difference in test performance was seen. The heterogeneity in test performances was magnified by the different proportions of pre- and postmenopause in the different study populations.

Sample size. Few studies discussed sample size issues, potentially leading to inappropriate conclusions, especially regarding comparability of test characteristics.

Failure to account for observer variability. No studies attempted to account for the effects of observer variation on the precision of estimates, although a few did calculate interobserver coefficients. For tests where the thresholds for normal and abnormal were based on either qualitative assessments (such as descriptions of ultrasound morphology) or quantitative measures (such as ultrasound morphology scores), this variability will have implications for the precision of sensitivity and specificity.

Prevalence and predictive value. We did not limit our analysis of test characteristics to studies from the United States. As the incidence of ovarian cancer is different in different countries, this influences the range of predictive values reported in the literature. Locations with low disease prevalence will have low PPVs compared with higher prevalence areas. The heterogeneity of study locations influenced the range of reported test characteristics and somewhat limits the comparability of the results.

Summary

Table 21 summarizes the pooled sensitivity and specificity estimates for CA-125 and the various imaging modalities.

Table 21. Pooled sensitivity and specificity estimates

Diagnostic Test	Pooled Sensitivity (95% CI)	Pooled Specificity (95% CI)
ULTRASOUND: MORPHOLOGY		
Scoring system: Sassone	0.86 (0.79 to 0.91)	0.77 (0.73 to 0.81)
Scoring system: DePriest	0.91 (0.84 to 0.95)	0.68 (0.49 to 0.82)
Scoring system: Ferrazzi	0.87 (0.80 to 0.92)	0.81 (0.62 to 0.91)
Scoring system: Finkler	0.82 (0.65 to 0.91)	0.78 (0.59 to 0.91)
Other	0.86 (0.82 to 0.89)	0.83 (0.76 to 0.88)
ULTRASOUND: DOPPLER		
Resistive index	0.72 (0.61 to 0.82)	0.90 (0.84 to 0.94)
Pulsatility index	0.80 (0.74 to 0.85)	0.73 (0.62 to 0.81)
Maximum systolic velocity	0.74 (0.56 to 0.86)	0.81 (0.59 to 0.83)
Presence of vessels	0.88 (0.80 to 0.92)	0.78 (0.65 to 0.87)
MORPHOLOGY PLUS DOPPLER	0.86 (0.79 to 0.91)	0.91 (0.80 to 0.97)
MRI	0.91 (0.86 to 0.94)	0.87 (0.83 to 0.90)
CT	0.90 (0.83 to 0.94)	0.75 (0.36 to 0.94)
FDG-PET	0.67 (0.52 to 0.79)	0.79 (0.70 to 0.85)
CA-125 (threshold > 35)	0.78 (0.75 to 0.81)	0.78 (0.71 to 0.82)

Abbreviations: CA-125 = cancer antigen 125; CI = confidence interval; CT = computed tomography; FDG = 18-Fluorodeoxyglucose; MRI = magnetic resonance imaging; PET = positron emission tomography

The use of established scoring systems in the evaluation of an adnexal mass by ultrasound morphology appears to perform slightly better than simple descriptive assessment. However, there does not appear to be a benefit of one scoring system over another. Based on small numbers of studies, 3D ultrasound shows some improvement over 2D. Although the pooled

sensitivity and specificity of MRI was the highest of any imaging modality, its performance was less consistent in studies where it was directly compared to other modalities such as CT and ultrasound.

Color Doppler assessment by RI, PI, and maximum systolic velocity are not superior to the more simple assessment of the presence or absence of arterial vessels within the mass. The efficacy of RI, PI, and maximum systolic velocity are hampered by the overlap in values of these measurements between benign and malignant masses.

Combined ultrasound morphology and color Doppler assessments have higher sensitivity and specificity compared to either alone. Although ultrasound morphologic evaluation by a gynecologist appears to be as reliable as that performed by a radiologist, there was no evidence of Doppler measurements done outside of the context of a radiology referral.

In postmenopausal women, an elevated CA-125 is useful for helping rule in ovarian cancer.

Qualitatively, there was a consistent trade-off across all tests between sensitivity and specificity.

The relatively low PPVs in all of the tests are particularly striking given that many of the included studies were done in preoperative patients; the likely "screening" done prior to a decision for surgery suggests that the PPV of a particular test in the initial evaluation of an adnexal mass is likely to be even lower.

Question 4: Explicit Scoring Systems

Question 4 is: What is the accuracy of explicit scoring systems which incorporate various combinations of imaging findings, patient risk factors, and/or CA-125 levels for detecting malignancy? Have these scoring systems been applied to a population of women before laparoscopy or laparotomy?

Approach

Explicit scoring systems were sought in the medical literature from among all studies of diagnostic assessment of adnexal or pelvic masses. We considered only scoring systems that combined data from more than one category of the following types of information: (1) imaging findings; (2) patient risk factors; and (3) laboratory data. Clinical prediction rules that utilized data entirely from only one category (for example, ultrasound based morphological indices[56]) are described as part of Question 3.

Imaging findings could include: (1) ultrasound based tests, such as transabdominal or transvaginal 2D ultrasound or Doppler ultrasound; (2) radiographic tests, such as CT; or (3) other imaging studies, such as MRI or PET scans.

Patient risk factors include menopausal status, age, or other risk factors.

Laboratory data was primarily CA-125, but we recorded data on other serum tumor markers as well.

Results of Literature Search and Screening

We identified 36 studies that met the inclusion criteria.[42,48,51-53,55,62,63,66,72,86,87,97,103,105,116,134,135,138,139,147,169,178,180,181,185,202-211] These are described in Evidence Table 4 (Appendix D*).

Study Characteristics

Scoring systems identified. The scoring systems were of several types. The most common were models developed using statistical modeling techniques such as logistic regression (or artificial neural networks) to develop estimates for predicted probability of malignancy. Such estimates were then used to construct clinical prediction rules (e.g., the Risk of Malignancy Index [RMI], which calculates a numeric score based on CA-125 level multiplied by a menopausal score and an ultrasound morphology score) and decision thresholds (e.g., for RMI, the most common threshold is 200). Other scoring systems used simple combinations of criteria based on individual modalities, which were then combined using Boolean *and* or *or* (e.g. CA-125 > 65 U/ml *and* ultrasound morphology score > 10 points). Some models were validated in separate populations from the data set used to develop the scoring systems either described as part of its initial development, or in subsequent publications by the original developers or others.

Types of data incorporated. The most common scoring systems used ultrasound, CA-125 and menopausal status. Some type of ultrasound data was used in all 36 publications; studies varied with regard to the type of ultrasound technology that was used. All used 2D ultrasound to evaluate morphology, some using transabdominal and many using transvaginal probes. Studies that used Doppler ultrasound used a variety of parameters, including measures as simple as detection of flow, or as complex as specific indices derived from Doppler-measured flow rates, such as the RI or PI. Many described scoring rules based on combinations of features of morphology (Finkler score) or combined morphology and blood flow.

CA-125 was a component of the scoring system in 30 reports; other serum tumor markers included CA-72-4, incorporated into two reports,[53,63] and the markers AFP, LDH, and hCG, were used in one report.[42] All studies that used these other serum markers also used CA-125.

Menopausal status was incorporated into scoring systems of 19 reports. The definition of menopausal status varied across studies, and in a few cases age was used as a proxy for clinically determined menopausal status. Three studies included only postmenopausal women,[62,63,135] and thus could not use this variable in the scoring system.

Physical examination was a component of scoring systems in six reports.[42,51-53,62,63]

Type of study populations. Most study populations were case series assembled at the time of referral for surgery and collected either at the point of preoperative ultrasound imaging or preoperative surgical evaluation. No studies were based in primary care clinical populations. One study described evaluation of adnexal masses detected during an ovarian cancer screening program.[51]

Reporting of study populations. Menopausal status of the study populations was described in 28 of the 36 reports; three reports included only postmenopausal women.[62,63,135]

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

Age was reported for the study population as a mean or median in 18 of 36 studies; it was reported in categories in one additional study. Symptom status was seldom described in the candidate reports.

Race/ethnicity was not reported in any of the studies.

Risk factors for ovarian cancer (besides menopausal status and age, describe above) were not reported, except in one study that reported the proportion of the study population that was nulliparous versus multiparous.[138]

Methodology. The methodological quality of the included studies may be described as follows:

Reference standard (handling of borderline). Some studies, particularly those assembled at the time of ultrasound investigation rather than surgery, encountered women with masses due to simple cysts with low risk of malignancy. Two studies allowed use of an operative report in lieu of histopathology as a reference standard,[87,116] and one used clinical followup without surgery as an alternate reference standard.[48]

Verification bias. Fourteen studies failed to verify disease status for all or a significant sample of test-negative women.

Test reliability. Only nine studies provided data on the reliability of test assessments.

Sample size. Only 11 of the reports described a priori recruitment targets or sample size calculations. We excluded studies with fewer than 50 women; however, some studies report subgroup analyses with fewer than 50 women, for example, the subset of postmenopausal women in Strigini et al.[169]

Use of appropriate statistical tests. The majority of reports (n = 28) used appropriate statistical analysis of the diagnostic data; however seven reports reported inadequate analyses.

Blinding. None of the reports described the use of techniques to blind investigators to the disease status of study patients.

Definition of positive and negative test. Most studies (n = 24) provided a priori definitions of a positive and negative test result; studies failed to meet this criterion most often when no explicit threshold was set a priori, but it was set based on study data.

Explicit validation method. Half of the reports (18/36) used some explicit validation method; many of the reports replicated previously described scoring systems in a new population. In many cases, these studies described new scoring systems which were not always validated.

The most common validation method was replication in a separate population. Two studies used validation techniques within a single study population: one split-sample,[209] and one bootstrap.[205]

Diagnostic Accuracy of Scoring Systems

This section considers the diagnostic accuracy of the RMI (Jacobs 1990) and subsequent replications and refinements (RMI2, RMI3, Jacobs 1993, and Timmerman models).

RMI. The first scoring system based on a statistical model was published in 1990;[116] it has been replicated in 11 subsequent clinical populations.[55,72,87,139,147,180,204,206-208,210] The diagnostic performance in these 12 studies is shown in Figure 17.

The RMI is a clinical prediction rule based on ultrasound, CA-125, and menopausal status data defined as follows:

RMI = U x M x CA-125

where ultrasound (transabdominal) is scored 1 point for each of the following characteristics: multilocular cyst, evidence of solid areas, evidence of metastases, presence of ascites, and bilateral lesions.

U = 0 for ultrasound score of 0
= 1 for ultrasound score of 1
= 3 for ultrasound score ≥ 2

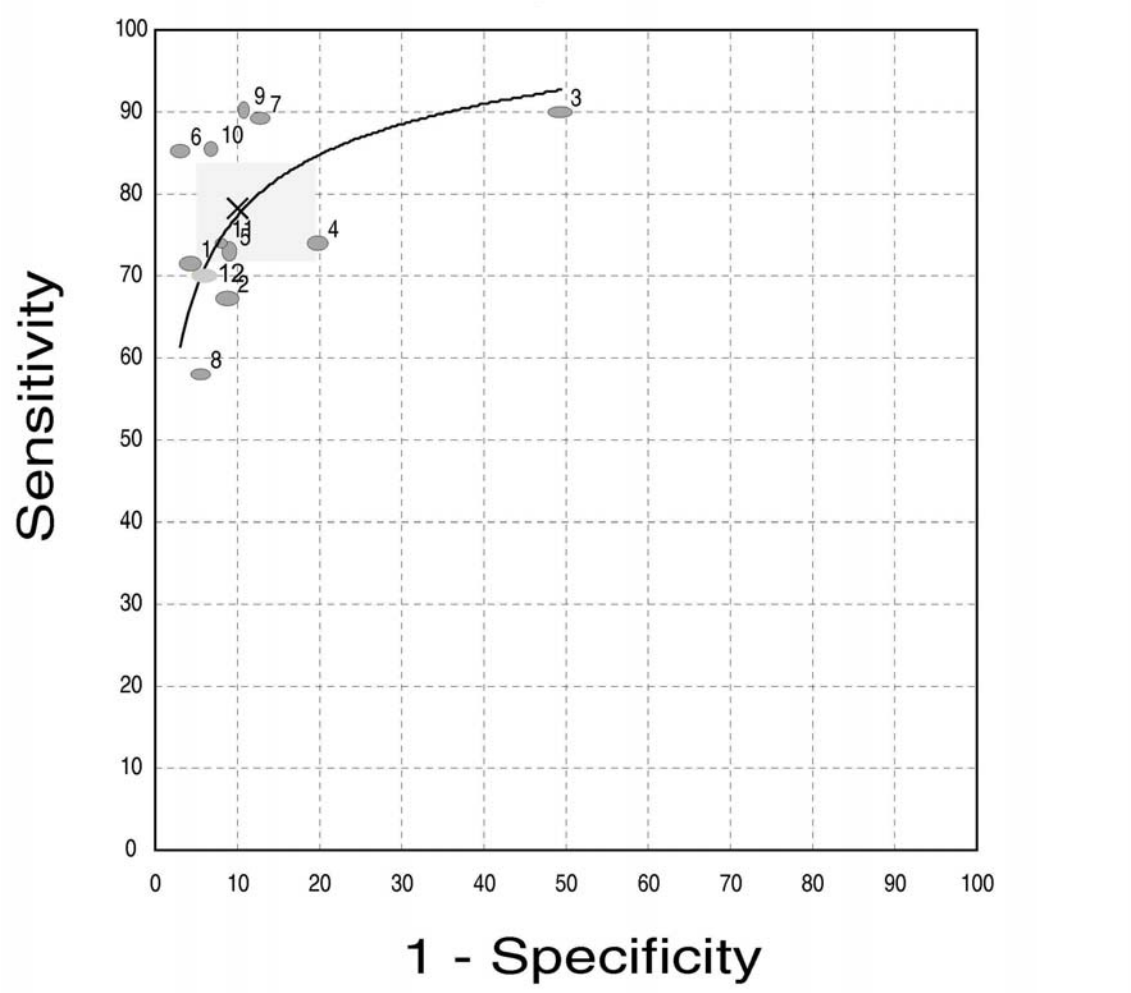
CA-125 = Serum CA-125 in U/ml

Menopausal status
M = 1 if premenopausal
= 3 if postmenopausal

In the initial report, Jacobs et al.[116] used the cutoff value of 200. At this cutpoint, sensitivity was 85 percent and specificity was 97 percent among a population of 143 women undergoing surgical investigation for an adnexal mass. The performance of the initial model (study 6 in Figure 17) has, in most studies, failed to be equaled in subsequent attempts at validation. Three of the subsequent 11 studies have similar performance (studies 7, 9, 10 in Figure 17).[72,87,208] It is notable that these three studies have fewer quality features (n ≤ 4) than the other eight studies (n ≥ 5 of 7 quality features).

When sensitivity and specificity are combined separately using a random-effects model, the pooled sensitivity is 0.78 (95% CI, 0.72 to 0.84) and the pooled specificity is 0.90 (0.81 to 0.95).

Figure 17. Performance of RMI model of Jacobs et al. (1990)[116] in development set (study 6) and subsequent validation studies using cutoff score of 200



Key to Figure 17: 1 = Tingulstad et al., 1996;[180] 2 = Timmerman et al., 1999;[210] 3 = Mol et al., 2001;[207] 4 = Lu et al., 2003;[206] 5 = Manjunath et al., 2001;[139] 6 = Jacobs et al., 1990;[116] 7 = Davies et al., 1993;[87] 8 = Morgante et al., 1999;[147] 9 = Obeidat et al., 2004;[208] 10 = Asif et al., 2004;[72] 11 = Aslam et al., 2000;[203] 12 = Dowd et al., 1993[55]

RMI2. In 1996, Tingulstad et al.[180] reported a refinement to the original RMI scoring system, commonly referred to as RMI2. RMI2 is defined identically to RMI except that new weights were used for the ultrasound and menopause components as follows:

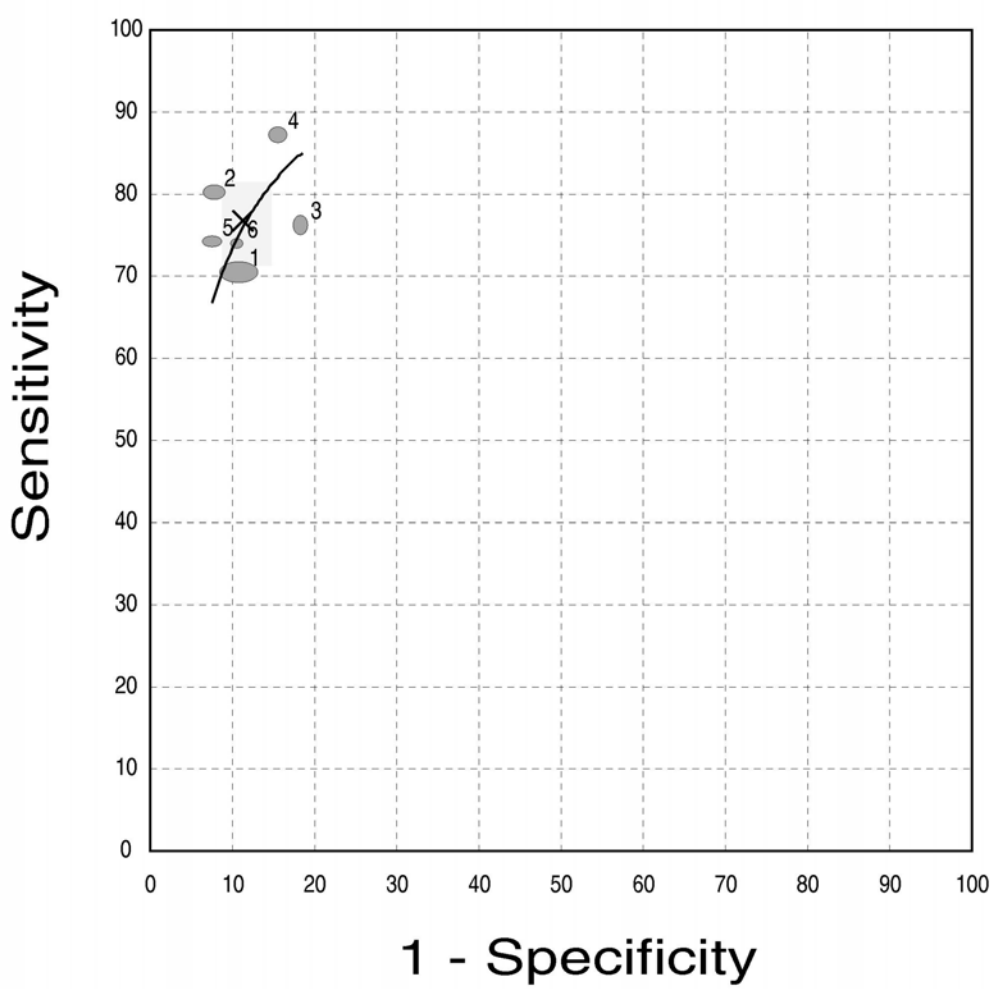
U = 1 for ultrasound score of 0-1
= 4 for ultrasound score ≥ 2

M = 1 if premenopausal
= 4 if postmenopausal

A cutoff value of 200 was also recommended for RMI2. Like the RMI, the RMI2 scoring system has been replicated.[134,139,147,207] The original report of RMI2 found sensitivity of 0.8 and specificity of 0.92. Subsequent validation studies have performed no better. These validation

studies all exhibited five or more quality features. The pooled sensitivity of all five studies is 0.77 (0.71 to 0.82), and pooled specificity 0.89 (0.85 to 0.91). The summary ROC curve is shown in Figure 18.

Figure 18. Performance of RMI2 model of Tingulstad et al. (1996)[180] in development set (2) and subsequent validation studies using cutoff score of 200



Key to Figure 18: 1 = Andersen et al., 2003;[202] 2 = Tingulstad et al., 1996;[180] 3 = Manjunath et al., 2001;[139] 4 = Ma et al., 2003;[134] 5 = Morgante et al., 1999;[147] 6 = Aslam et al., 2000[203]

RMI3. Subsequently, a further refinement to the RMI and RMI2 was reported by Tingulstad et al.[211] This third scoring system is defined identically to RMI and RMI2 except that new weights were used for the ultrasound and menopause components as follows:

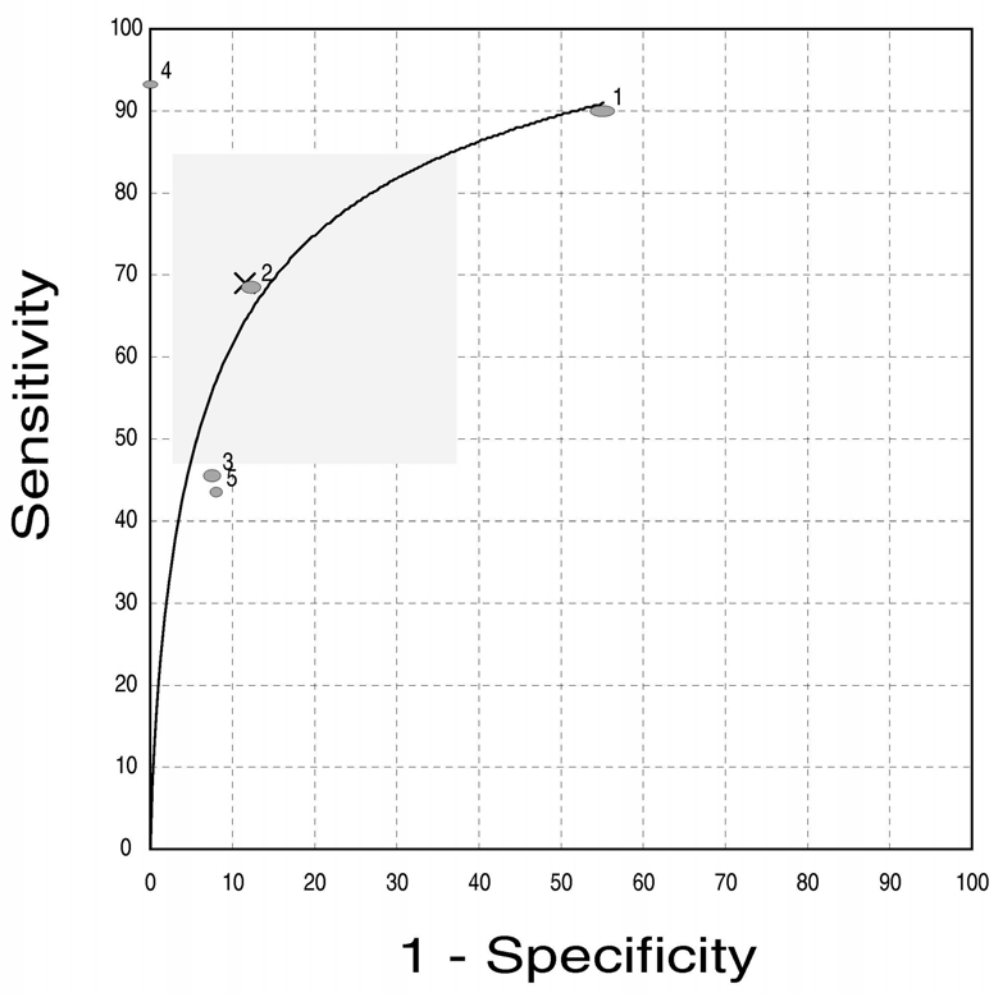
U = 1 for ultrasound score of 0-1
= 3 for ultrasound score $\geq$ 2

M = 1 if premenopausal
= 3 if postmenopausal

A cutoff value of 200 was also recommended for RMI3. The RMI3 scoring system has been replicated in one additional study.[139] The original report of RMI3 found sensitivity of 0.71 and specificity of 0.92, while the validation study reported very similar performance, with sensitivity of 0.74 (0.65 to 0.83) and specificity of 0.91 (0.83 to 0.99).

Tailor and subsequent replications. Tailor et al.[209] reported a scoring system based on an artificial neural network method that was based on a small population of 67 women total, 15 of whom had malignancies. Unlike the RMI family of systems described above, this system did not include CA-125, but considered age, menopausal status, and a variety of ultrasound morphological features and Doppler indices. While this system reported using 52 cases as a training set and 15 cases as a test set, the performance of the system was reported only for the study population as a whole: sensitivity 0.93 (95% CI, 0.81 to 1.0) and specificity 1.0 (0.94 to 1.0). Subsequently four studies have replicated this system showing markedly poorer diagnostic performance (Figure 19) when applied to separate populations, consistent with over-fitting in the initial model development.[185,203,204,207]

Figure 19. Performance of model of Tailor et al. (1999)[209] in development set (4) and subsequent validation studies



Key to Figure 19: 1 = Mol et al., 2001;[207] 2 = Valentin et al., 2001;[185] 3 = Aslam et al., 2000;[203] 4 = Tailor et al., 1999;[209] 5 = Aslam et al., 2000[204]

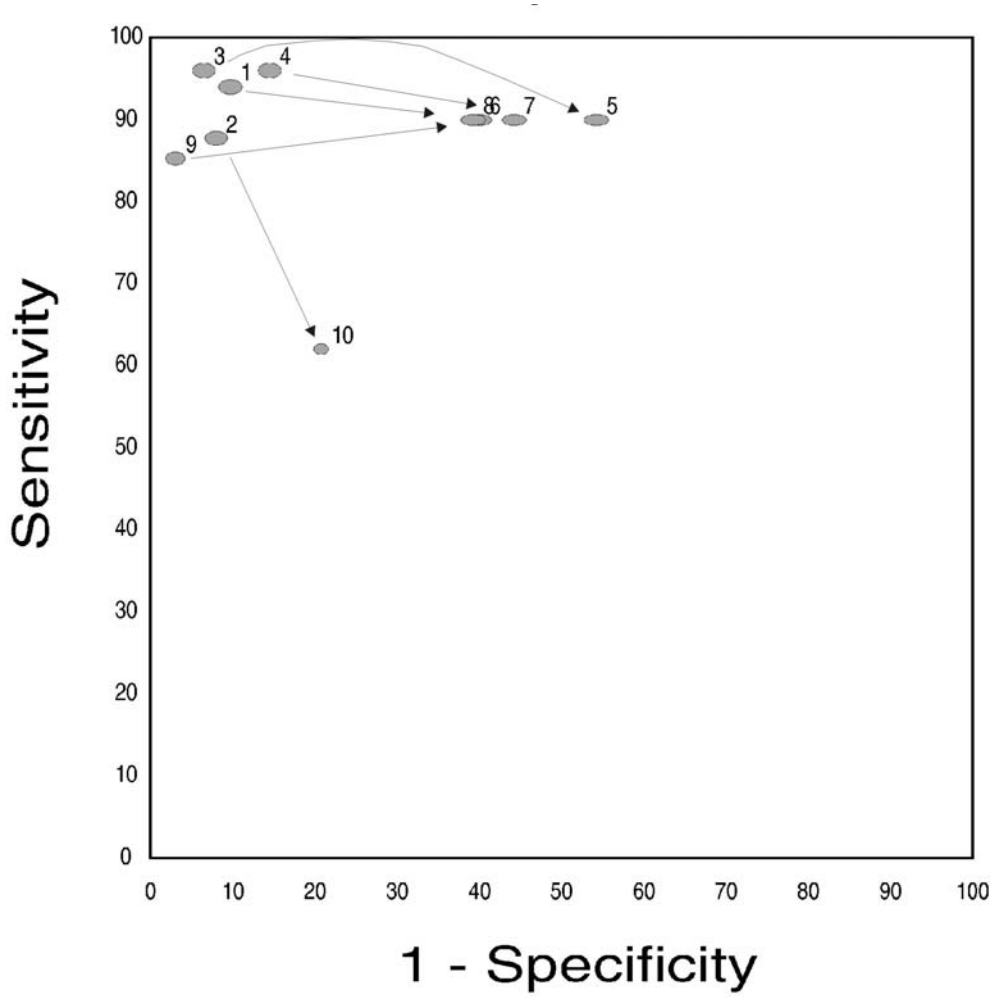
Twenty other scoring systems have been described, none of which has been as extensively replicated as the systems described above. Five of these other scoring systems have been validated in one other population as shown in Table 22; each of the systems was based on ultrasound morphology, CA-125, Doppler, and menopausal status. The models were: Timmerman LR1,[178,210] Timmerman AAN1,[178,207] Timmerman AAN2,[178,207] Timmerman LR2,[178,207] and Jacobs 1993.[207,212]

Table 22. Performance of other scoring systems at initial derivation and subsequent replication in another population

Initial description	Subsequent validation	Sensitivity (95% CI)		Specificity (95% CI)	
		Initial development	Replication	Initial estimate	Replication
Timmerman LR1[210]	Valentin 2001[185]	0.87 (0.79 to 0.97)	0.62 (0.44 to 0.80)	0.92 (0.87 to 0.97)	0.79 (0.68 to 0.90)
Timmerman AAN1[178]	Mol et al. 2001[207]	0.94 (0.87 to 1.0)	0.90 (0.79 to 1.0)	0.90 (0.85 to 0.96)	0.60 (0.52 to 0.68)
Timmerman AAN2[178]	Mol et al. 2001[207]	0.96 (0.90 to 1.0) (0.91	0.90 (0.79 to 1.0)	0.94 (0.89 to 0.98)	0.46 (0.38 to 0.54)
Timmerman LR2[178]	Mol et al. 2001[207]	0.96 (0.90 to 1.0)	0.90 (0.79 to 1.0)	0.86 (0.79 to 0.92)	0.56 (0.48 to 0.64)
Jacobs 1993[212]	Mol et al. 2001[207]	0.85 (0.74 to 0.96)	0.90 (0.79 to 1.0)	0.97 (0.94 to 1.0)	0.61 (0.53 to 0.69)

In each case, the initial diagnostic performance described by the system significantly degrades on replication in another population (Figure 20).

Figure 20. Performance of various other scoring systems in development and validation studies in separate populations

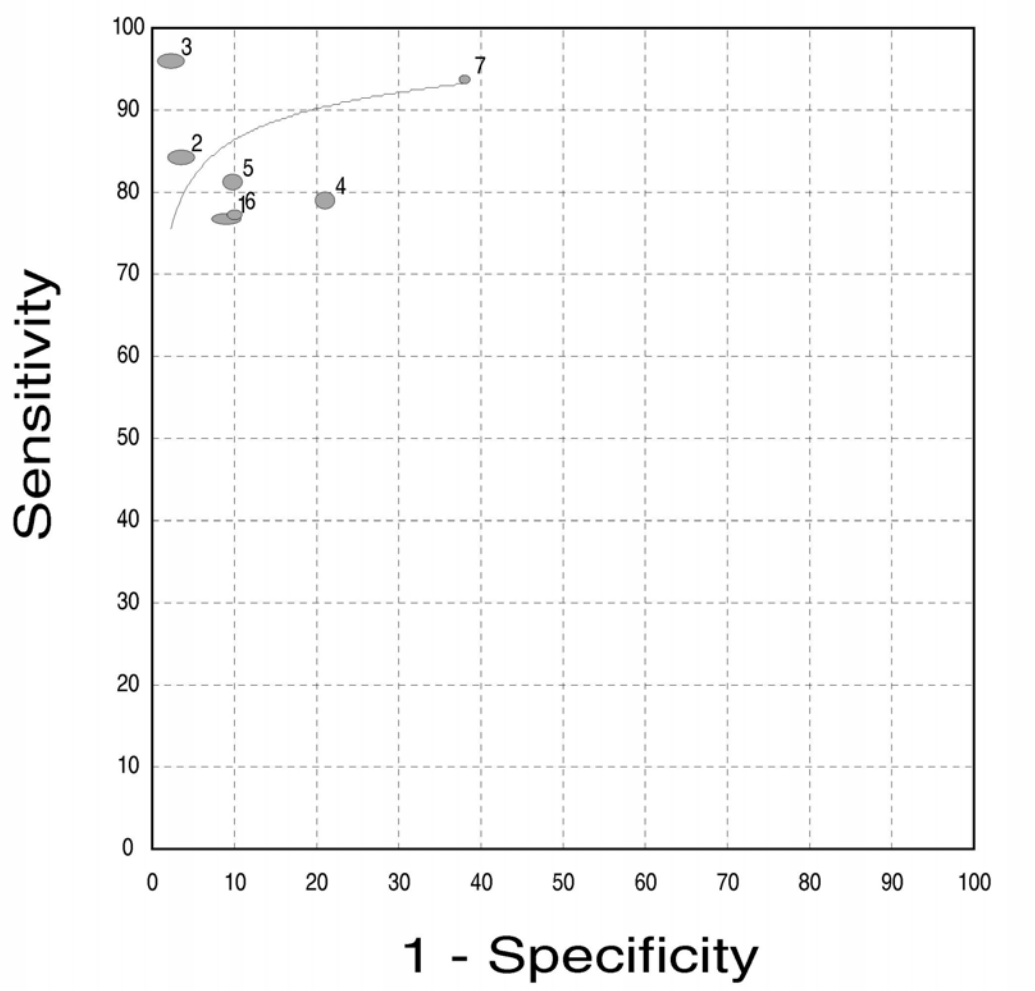


Arrows indicate change in performance estimate from development (start of arrow) to validation (end of arrow) for paired studies of each scoring system.

Key to Figure 20: 1-4 = Timmerman et al., 1999;[210] 5-8 = Mol et al., 2001;[207] 9 = Jacobs et al., 1993;[212] 10 = Valentin et al., 2001[185]

Ten additional systems were described in seven reports.[42,48,53,63,181,203,205] Most of these studies used logistic regression or artificial neural network modeling methods to derive a new model. One used bootstrap validation techniques,[205] but none was validated in another study population. One of these studies[203] reported on newly fitted logistic regression models created by forcing variables that were include in previously described scoring systems.[178,209,213] Aslam et al.[204] constructed three separate models based on each possible pairwise combination of the three previously described models. The diagnostic performance of these miscellaneous unvalidated models is shown in Figure 21.

Figure 21. Performance of various other unvalidated models



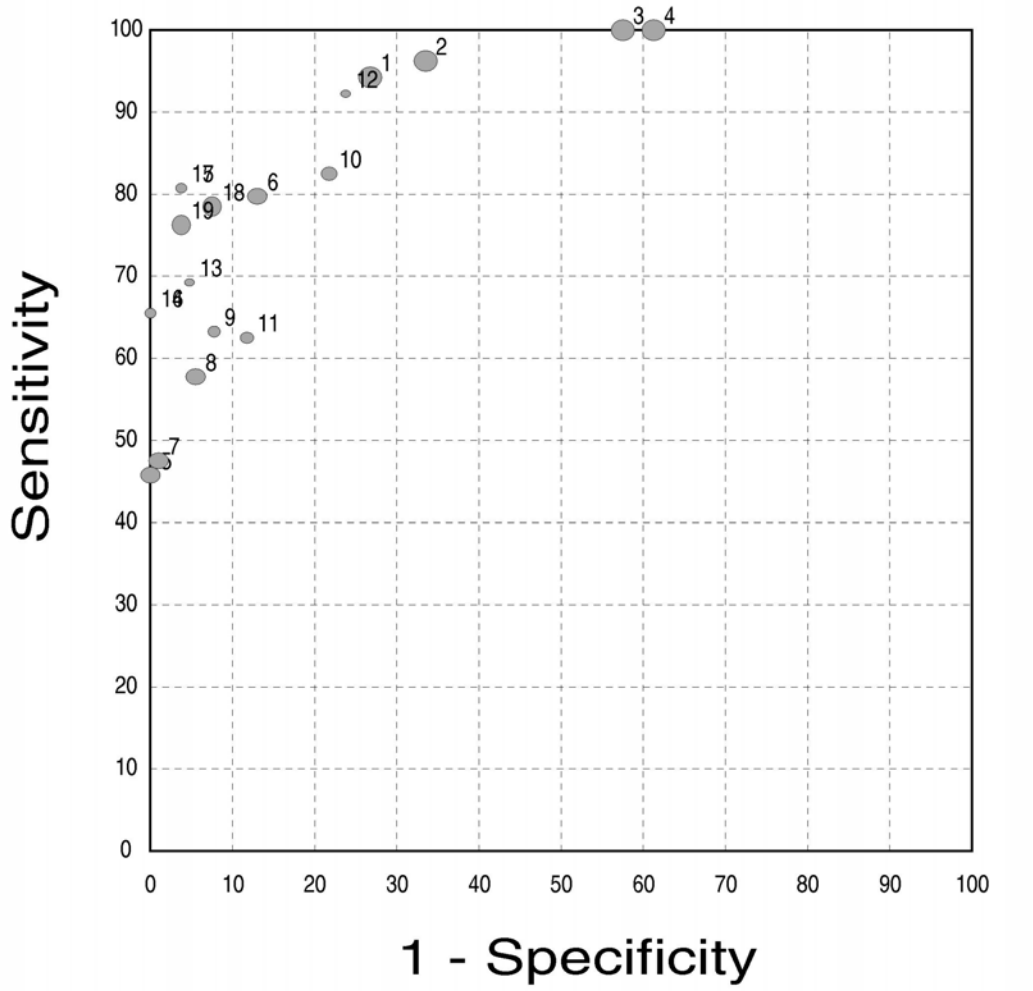
Key to Figure 21: 1 = Twickler et al., 1999;[48] 2-3 = Biagiotti et al., 1999;[205] 4 = Torres et al., 2002;[181] 5 = Schutter et al., 1998;[63] 6 = Balbi et al., 2001;[53] 7 = Roman et al., 1997[42]

Thirteen further reports describe the diagnostic performance of simple rules for combining single test or single modalities into a decision rule.[42,51,52,62,63,66,86,97,103,105,135,138,169] None of these criteria has been validated in another population. Each of these studies used dichotomous rules for two or more tests (or modalities) and then combined them using a simple rule like "malignant if any test positive" (Boolean *or*) or "malignant if all tests positive" (Boolean *and*). Some of the studies reported diagnostic performance of several different simple rules.

Twelve of these studies used ultrasound and CA-125, five incorporated physical exam, two included other serum tumor markers[42,63] and one used age over 50 years.[138]

Six of these studies reported results for postmenopausal women separately: in three studies, the entire study population was postmenopausal[62,63,135] while three studies reported diagnostic performance for the postmenopausal subgroup separately.[103,105,169] The diagnostic performance of 18 simple combination rules in these six studies is shown in Figure 22.

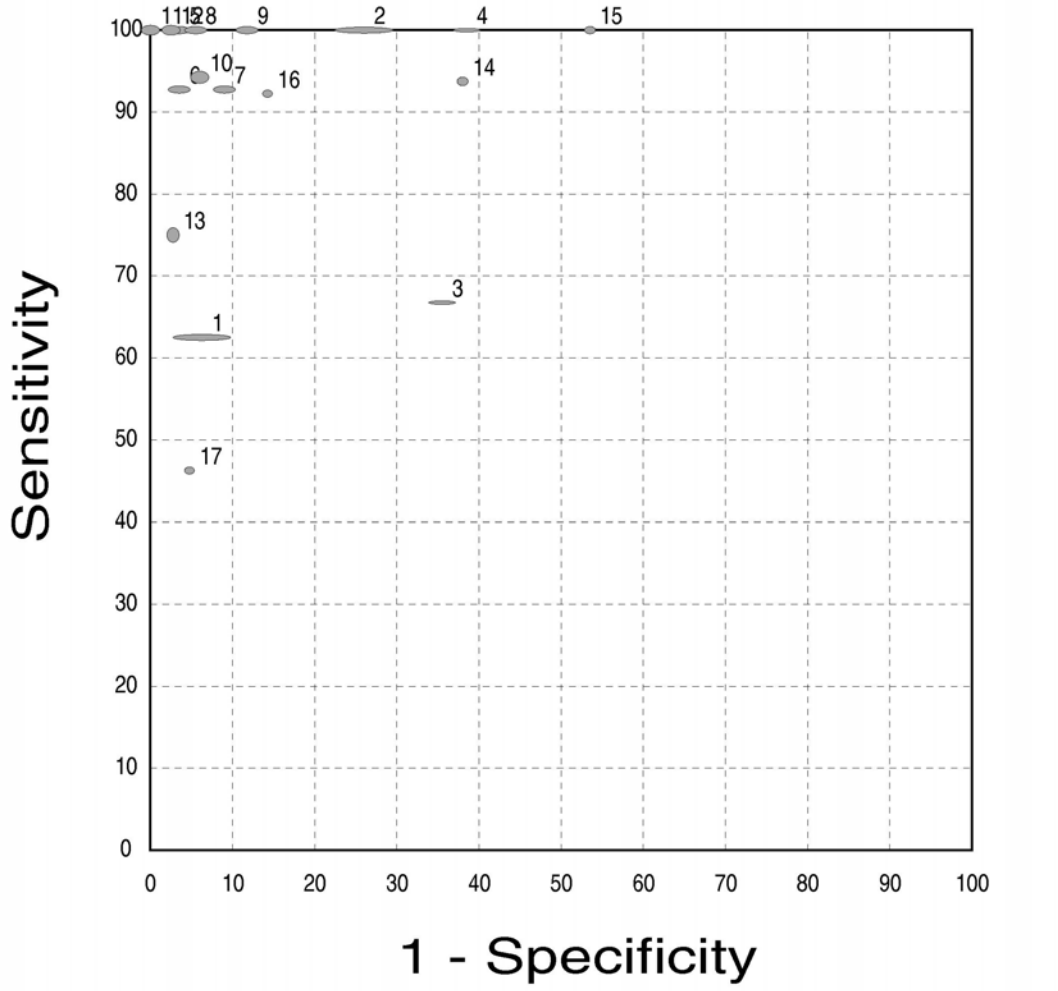
Figure 22. Performance of unvalidated simple combination rules in postmenopausal women only



Key to Figure 22: 1-4 = Maggino et al., 1994;[135] 5-8 = Schutter et al., 1998;[63] 9-11 = Schutter et al., 1994;[62] 12-13 = Strigini et al., 1996;[169] 14-17 = Guerriero et al., 1998;[103] 18-19 = Guerriero et al., 2002[105]

In contrast, the diagnostic performances of 17 simple combination rules in studies that include both premenopausal and postmenopausal women in the study population are shown in Figure 23.

Figure 23. Performance of unvalidated simple combination rules in mixed pre- and postmenopausal populations



Key to Figure 23: 1-2 = Andolf et al., 1990;[52] 3-4 = Adonakis et al., 1996;[51] 5-9 = Mancuso et al., 2004;[138] 10 = Gadducci et al., 1988;[97] 11-12 = Chou et al., 1994;[86] 13 = Alcazar et al., 1999;[66] 14-15 = Roman et al., 1997;[42] 16-17 = Strigini et al., 1996[169]

The results show a wide range of sensitivity and specificity. This variation reflects differences in decision thresholds (e.g., CA-125 > 35 U/ml versus CA-125 > 65 U/ml) and in the rules for combining tests (e.g., use of Boolean *or* versus *and* when combining results of two or more tests).

Discussion

No scoring systems were both developed and validated expressly for evaluating adnexal masses in postmenopausal women. Existing scoring systems that have been validated have all been developed in mixed pre- and postmenopausal populations. Those scoring systems that have been described in populations of postmenopausal women were neither rigorously developed (they consist of simple combination rules) nor validated in other populations.

The highest demonstrated specificity obtained with these scoring systems appears to be in the range of 90 to 95 percent and, at this range of specificity, the sensitivity appears to be in the range of 65 to 80 percent. However, as suggested by the performance in the few populations of postmenopausal women studied, the same degree of sensitivity and specificity is unlikely to be possible. Reliable estimates of the diagnostic performance of scoring systems cannot be determined from these studies.

This review of scoring systems demonstrates several important limitations of predictive models and has important implications for the clinical usefulness of these models and the future research in this area of inquiry. First, validation in an external population is critical to obtain accurate estimates of diagnostic performance, because all modeling techniques lead to overestimation of diagnostic performance in the data from which it was derived. This overestimation of diagnostic performance is clearly demonstrated by comparing the development and validation studies described for RMI, Tailor, and other scoring systems (Figures 17-20). The studies described here suffer from being relatively small for modeling; reliable variable selection and parameter estimation requires at least 10 to 15 cases (in this case, ovarian malignancies) for every term selected in a predictive model. Few, if any, met this statistical rule of thumb. This limitation is particularly apparent in the case of the Tailor model, where subsequent studies demonstrated a high degree of overestimation of the original model. Third, these studies used populations that were identified following referral for surgery in most cases, after some filtering had already occurred. Furthermore, these studies failed to describe the initial presentation (symptomatic or asymptomatic, palpable or non-palpable mass) of women eventually enrolled. Thus, the applicability of these studies to women in primary care, where an adnexal mass is often first noted, is uncertain.

Question 5: Monitoring Women with Suspected Benign Masses

Question 5 is: Among women with suspected benign masses on initial investigation, what are the sensitivity and specificity of monitoring with periodic CA-125 and/or interval ultrasound examinations for detecting malignant masses? How does the interval of testing/definition of change affect sensitivity and predictive value?

Approach

For each study we sought to identify a population of patients with a screening abnormality which was "probably benign" and which the authors felt did not meet criteria for immediate surgical intervention. We then attempted to define the outcomes of further testing in the defined population, including the results of subsequent testing and final clinical outcome as defined by a pathology report or extended clinical followup. The interpretation of results is limited by the narrow scope of Question 5. Specifically, it is often difficult to identify a subgroup of patients with a screening abnormality which could be defined as a "suspected benign lesion" within larger screening studies. Often, results are not stratified with respect to these sub-populations, making it difficult to calculate sensitivity and specificity of the followup testing. In addition, by definition, it is also difficult to estimate the "sensitivity" of a followup regimen. We assumed

that this refers to detection of cancer as part of the followup regimen, and that women with cancer diagnosed outside of the followup were "false negatives."

Results

We identified nine articles meeting the criteria for this question;[40,127,135,145,214-218] these are summarized in Table 23, with details in Evidence Table 5 (Appendix D*). Five were population-based screening studies of asymptomatic, postmenopausal patients without known ovarian masses;[40,127,145,214,217] one was a voluntary screening program.[216] All addressed to some degree the use of interval ultrasound for detecting malignant masses. Although several used CA-125 as part of their followup, none reported any results based on the use of interval CA-125 in a population with adnexal lesions. None addressed the effects of changing the interval of testing on sensitivity and predictive value; the disparate nature of the studies prohibited any inferences on the effect of test interval on sensitivity.

Table 23. Studies of followup regimens for benign-appearing lesions

Study	Population	N	Followup interval	Length of followup	Loss to followup	True/false positives detected during followup	Cancers missed
Population-based studies (followup of "benign" masses identified in screening)							
Menon et al., 200[145]	Followup of scans considered "equivocal"	17	"Equivocal" scans followed every 6 weeks until clearly normal or abnormal; normal scans followed with CA-125 every 3 months	Median 6.8 years	Not reported	1 cancer/5 benign lesions	0 (1 within 6 weeks of initial test, before first followup scan)
Modesitt et al., 2003[40]	Followup of simple cysts < 10 cm	2,763	TVUS every 3-6 months for simple cysts	Mean 6.3 years	Not reported	7 cancers/0 benign lesions	3 cancers, none developed in the original cyst
Schin-caglia et al., 1994[217]	Followup of post-menopausal ovaries > 9 cc, or with simple cyst	347	If cyst: followed with ultrasound every 6 months; if change, referred; others: referral if unchanged at 3 and 6 months	"At least" 1 year	Not reported, but all had "at least 1 year"	2 cancers/96 benign lesions	None in 249 not referred
Kurjak et al., 1994[127]	Followup of post-menopausal women with simple cyst > 2.5 cm but < 5 cm, resistive	88 (of 404 with sim-ple cysts)	Repeat ultrasound every 6 months	6 months	Not reported	1/17 with benign lesions	0

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

Study	Population	N	Followup interval	Length of followup	Loss to followup	True/false positives detected during followup	Cancers missed
	index ≥ 4.1)						
Castillo et al., 2004[214]	Followup of post-menopausal women with simple cyst < 10 cm	215	Repeat ultrasound and CA-125 in 3 months, then every 6 months	Median 27 months	30.6%	0/44 benign masses	1
Case series (clinical history prior to identification of mass not routinely described)							
Valentin and Akrawi, 2002[218]	Followup of post-menopausal women with low score on ultrasound malignancy risk scale	162	Repeat ultrasound 3, 6, 9, and 12 months, then every 12 months; test positive if increase in size or cyst more complex	Median 3 years	0 (cancer and mortality tracked through registry)	0 cancers/7 patients underwent surgery for change	0
Mag-gino et al., 1994[135]	Followup of post-menopausal women with cysts < 5 cm, thin wall, no septae, no free fluid	45	Details on followup strategy not reported	Not reported	4.4%	0/0	0
Levine et al., 1992[216]	Followup of voluntary screening of post-menopausal women with unilocular simple cyst	32	Repeat ultrasound every 3 months x 1 year, then every 6 months	"Over half at least one year"	22.2%	0/0	0
Gold-stein et al., 1989[215]	Followup of post-menopausal women with simple cysts ≤ 5 cm	16	Repeat ultrasound (abdominal)	Mean 29 months	6 (12% of original 48)	0/2 with benign lesions	0

Abbreviations: CA-125 = cancer antigen 125; TVUS = transvaginal ultrasound

Menon et al.[145] performed a large prospective screening study of 22,000 postmenopausal women older than 45 years. Initial screening consisted of CA-125; patients with CA-125 ≥ 30 underwent endovaginal ultrasound evaluation. Results were interpreted as normal (ovarian volume < 8.8 ml/normal morphology), equivocal (volume < 8.8 ml, abnormal morphology), or abnormal (volume ≥ 8.8 ml). Normal morphology was defined as uniform hypoechogenicity and smooth outline. Abnormal morphology was defined as simple cysts or complex lesions. Patients with normal scans were triaged to repeat CA-125 every 3 months for a year and subsequently returned to yearly screening; median followup was 6.8 years, with loss to followup not reported. Patients with abnormal scans were referred to a gynecologist for consideration of surgical intervention. Patients with equivocal scans were triaged to repeat ultrasound at 6-week intervals until a scan could be classified as normal or abnormal. Of 741 patients who were triaged to

ultrasound, 20 (2.7 percent) index cancers were identified. We focused on the group of patients with "equivocal" scans who were triaged to interval testing in an attempt to answer the study question. There were 17 equivocal scans. Of these, nine had simple cysts which were followed and did not result in a cancer diagnosis (true negatives). One patient died of pneumonia prior to her first repeat ultrasound, and one died of advanced ovarian cancer prior to her first repeat ultrasound; this cancer death could possibly be considered a false negative for the followup strategy, although it could also be considered a false negative from the original study since the death occurred within 6 weeks of the initial scan. Six patients were scheduled for surgery following an equivocal scan, presumably due to abnormal followup ultrasound. One of these had ovarian cancer (true positive), and the other five had benign disease (false positive). Because the number of equivocal scans was so small, and because the classification "equivocal" does not necessarily imply that the lesions were felt to be "suspected benign" as designated in Question 5, it is not possible to calculate the sensitivity and specificity of prolonged monitoring strategies using this study. The authors do not draw any conclusions regarding the appropriateness of interval testing.

Modesitt et al.[40] performed a large screening study of 15,106 asymptomatic women at least 50 years old without a history of ovarian cancer. Patients were screened with TVUS. Criteria for abnormality were ovarian volume > 10 ml and any morphologic abnormality, including simple or complex cysts. Patients with abnormal TVUS were triaged to repeat TVUS in 4 to 6 weeks, with Doppler flow ultrasound, CA-125 level, and tumor morphology indexing performed at the second visit. Patients with simple unilocular cysts which were considered likely benign were triaged to repeat TVUS every 3 to 6 months. Mean followup was 6.3 years. Two thousand and seven hundred and sixty-three (2,763) women were diagnosed with 3,259 unilocular cysts. Spontaneous resolution of unilocular cysts occurred in 2,261 (69.4 percent) of lesions. Ten patients subsequently developed ovarian cancer. Seven of these had additional abnormal areas which subsequently developed on TVUS (considered true positives because they were subsequently identified by interval testing). Two developed ovarian cancer after the cyst in question had resolved on sonographic followup (these might be considered false negatives). One patient developed cancer in the ovary opposite the cyst being followed (this might also be considered a false negative). Calculated on a per-patient basis, the sensitivity and specificity of followup testing in the population with a simple unilocular ovarian cyst are 70 percent (95% CI, 41.6 to 98.4 percent) and 100 percent (99.9 to 100 percent), respectively. Because none of the unilocular cysts subsequently developed into a cancer, the sensitivity and specificity improve to 100 percent (57.1 to 100 percent) and 100 percent (99.9 to100 percent), respectively, when calculated on a per-lesion basis. Followup time is a major strength of this study. The authors conclude that unilocular ovarian cysts are associated with a very low risk of malignancy and can be safely followed with serial ultrasound.

Schincaglia et al.[217] performed a screening study of 3,541 asymptomatic postmenopausal patients. All patients underwent transabdominal ultrasonography with assessment of ovarian volume and morphology. Patients were divided into four groups based on the results of the initial ultrasound. All patients with ovarian volume > 15 ml (Group 4) were referred for repeat "level II" ultrasonography for morphologic assessment and fine needle aspiration (FNA) when feasible. Patients with ovarian volume between 9 and 15 ml (Group 3) were triaged to followup ultrasound at 3 and 6 months. Patients with ovarian volume < 9 cm but a cystic appearance (Group 2) were triaged to followup ultrasound in 6 months. Patients with ovarian volume < 9 ml and homogeneous appearance (Group 1) were considered negative and had no further

intervention. Clinical followup at 1 year and pathology results if surgery was performed were considered the reference standard. Two hundred and eighty-three (283) patients (Groups 2 and 3) were deemed appropriate for followup using repeat ultrasound at 3- to 6-month intervals without the need for immediate referral for FNA/surgery. Of these 283 patients, 34 subsequently developed concerning ultrasound findings and were referred for a level II scan and/or possible FNA. The clinical results of this group of 34 are not given separately. Of the 249 who had non-concerning followup scans, none developed cancer with followup of at least 1 year ("true negatives"). Therefore, the specificity of ultrasound followup is 100 percent (95% CI, 98.8 to 100 percent) for patients with an initial abnormal but "probably benign" ultrasound. Sensitivity within this group cannot be calculated with the information given in the publication. The ability to answer Question 5 would be enhanced if specific outcomes of each of the four groups defined by the authors had been given. The study was also limited by the fairly short followup interval and the lack of prior or concurrent validation of the ultrasonographic groups defined in the study.

Kurjak et al.[127] screened 5,013 women 40 years old or older (30.6% postmenopausal), of whom 404 had simple cysts with a diameter between 2.5 and 5 cm and a resistive index greater or equal to 0.41. These women received a followup scan in 6 months. Investigators reported the results of 88 women for whom the 6-month scan results were available. The definition of change prompting further diagnosis was not explicitly described. Of the 88 women, 18 ultimately underwent surgery based on the findings at 6 months, with one cancer detected and 17 benign lesions. Results stratified by menopausal status were not provided. This study was limited by lack of details on clinical decision rules, and short followup.

Castillo et al.[214] screened 8,794 postmenopausal women; 215 had simple unilocular cysts less than 10 cm in diameter. Twelve percent of these masses were asymptomatic. These women underwent repeat ultrasound and CA-125 in 3 months, with subsequent followup studies every 6 months. Progression was defined as an increase in diameter of 1 cm or more, regression as a decrease of 1 cm, and resolution as absence of the cyst at 2 consecutive visits 6 months apart. Median followup was 27 months. There was one interval ovarian cancer between studies, and 44 women had benign masses removed. Although this study was among the highest quality studies in terms of reporting of relevant data, it is limited by the relatively small size and the high loss to followup (30.6%).

The remaining four studies[135,215,216,218] were all small (less than 200 patients), and of variable quality (Table 23). None reported any interval cancers in patients receiving followup, or cancers detected during followup. The study of Valentin et al.[218] was notable for length of followup (median 3 years) and complete ascertainment of followup status using Swedish cancer and death registry data.

Discussion

There are limited data available to support a global definition of "probably benign" ovarian lesions or to support a specific method of interval testing to identify ovarian malignancy among patients in whom such lesions have been identified. For the most part, studies are limited by small size, variable length of followup, variable definitions of significant change and thresholds for intervention, and methods for followup.

The question of how best to define and evaluate "sensitivity" of followup regimens is a difficult one. Several factors need to be considered. First, interval cancers presenting between the initial study and the first followup visit may well be considered false negatives of the initial

study; alternatively, they may reflect a too-long followup interval. Second, given the lack of data on the natural history of ovarian cancer, it is unclear whether cancers developing in benign-appearing lesions represent subclinical cancers present at the time of the initial diagnosis, or new cancers representing malignant transformation of a benign cyst. If the latter, then the ultimate success of any followup regimen may depend as much on the natural history of a given malignancy as on the sensitivity and specificity of the tests used for followup. Finally, cancers identified during followup should ideally have high survival rates (although whether such high survival rates would reflect the efficacy of the followup or the biology of cancers which are associated with benign-appearing cysts is unclear). The number of cancers identified in the reviewed studies was too small to draw any inferences about relative survival.

Overall, only two interval cancers occurred during followup in the studies identified (one prior to the first followup scan), and 10 cancers were identified during followup. As noted, an additional three cancers developed after resolution of a cyst or in the contralateral ovary. The highest quality study[40] provides good evidence for the safety of prolonged followup with interval TVUS at 3- to 6-month intervals for patients with unilocular ovarian cysts of up to 10 cm in diameter, and the findings of the other studies are consistent with this conclusion.

Question 6: Surgical Morbidity and Mortality

Question 6 is: Among women with adnexal masses, what are the morbidity and mortality from diagnostic surgery (laparoscopy or laparotomy)? At what point does the risk of surgery outweigh the risk of detecting malignancy?

Approach

We searched the literature for studies that reported the morbidity and mortality of surgical management of adnexal masses. We also used the Nationwide Inpatient Sample (NIS) discharge database, maintained by the Agency for Healthcare Research and Quality (AHRQ), to obtain estimates of morbidity and mortality associated with diagnostic laparoscopy or exploratory laparotomy for a range of diagnoses associated with adnexal masses. The NIS is limited to inpatient procedures and does not cover ambulatory surgical centers, where some adnexal masses are likely to be managed, especially those masses thought to have a low likelihood of cancer. In addition to surgical complications, we also examined articles that provided data on the test characteristics of frozen section pathologic diagnosis; especially in the setting of minimally invasive procedures, false negative results on frozen section might lead to suboptimal surgical management and delayed therapy, while false positive results might lead to more extensive surgery than necessary, with possible implications for increased surgical morbidity and affects on ovarian function.

Results of Literature Search and Screening

We identified 24 articles that met our inclusion criteria;[32,37,41,219-239] these are summarized in Evidence Table 6 (Appendix D*). Twenty-two articles reported on the morbidity and mortality of surgical management of adnexal masses.[32,37,41,219-234,237-239] In addition, two of the included articles reported on the sensitivity and specificity of frozen section;[220,236] false negative frozen section results could lead to inadequate surgical management and delayed treatment, while false positive results could lead to more extensive surgery than necessary. Finally, one of the included articles addressed the potential effect of conservative surgery for removal of an ovarian cyst resulting from endometriosis (endometrioma) on subsequent fertility.[235]

Methodological Quality of Included Studies

Size of population. None of the papers provided a description of the referral base; two[32,37] were limited to gynecologic oncology practices. Lack of information on the referral base prevents assessment of generalizability. Since all of these studies were performed in centers experienced in laparoscopic surgery, the generalizability may well be limited.

Number of cases. Five studies had fewer than 200 cases, with correspondingly wide confidence intervals for reported event rates. Two studies had larger numbers of cases, 683[230] and 757.[219] However, the study by Marana et al.[230] was limited to women under 40.

Patient selection. None of the studies reported how patients were referred to the surgical practices. All provided criteria for laparoscopic management of masses, based on various criteria to suggest high or low risk of malignancy. We found two trials where patients were randomized to laparoscopy or laparotomy,[224,225] but randomization methods were not well described.

Application of reference standard. In this sense, "reference standard" refers to the method by which a complication was diagnosed. Only two studies described followup beyond 8 weeks, but they did not detail whether all patients underwent similar followup protocols.

Results

There were three deaths in one study of 146 patients (all undergoing laparoscopy), and none in any of the other studies (a total of 5,599 patients). Pooling all patients, the mortality was 0.05 percent, with a 95% CI of 0.01 to 0.17 percent.

Table 24 shows the results from individual studies. The two randomized studies[224,225] both showed lower morbidity with laparoscopy compared to laparotomy, although only one of them[224] had sufficient power to show a statistically significant difference. Although the study of Deckardt and colleagues[224] was randomized, there were substantial differences in the procedures performed in each arm. Laparoscopy patients tended to undergo more conservative procedures: they were significantly more likely to have cystectomy (60.0 vs. 20.2 percent), less likely to have oophorectomy (0.8 vs. 20.2 percent), and less likely to have bilateral salphingo-oophorectomy (4.0 vs. 21.4 percent). Both studies where laparoscopy was directly compared to laparotomy showed increased complication rates (primarily postoperative complications) among the

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

laparotomy patients. The four non-randomized studies all showed higher morbidity rates with laparoscopy, but there were substantial differences in patient selection criteria.

In series of laparoscopy cases, morbidity rates ranged from 0.9 percent to 22.1 percent (Table 24); series differed widely in their selection criteria for laparoscopic management of the mass. Few stratified results based on menopausal status; in some cases, postmenopausal patients were explicitly excluded. In one study where multivariate analysis was performed to assess for risks of morbidity, performance of additional procedures (hysterectomy) significantly increased the risk of morbidity, while a history of hysterectomy increased the likelihood of conversion to laparotomy (presumably because of increased technical difficulty secondary to postoperative adhesions).[226]

Table 24. Morbidity in series of patients undergoing surgical management of adnexal masses

Study	N	Patient population	Complication rate (95% CI)	Notes
Randomized trials of laparoscopy versus laparotomy				
Deckardt et al., 1994[224]	192	22.4% laparoscopy, 26.4% laparotomy postmenopausal	Laparotomy: 30.3% (21.8 to 42.3%) Laparoscopy: 11.2% (6.8 to 18.7%)	"Randomized," but some differences between two arms 3.5% conversion
Fanfani et al., 2004[225]	100	Laparoscopy: 10% postmenopausal Laparotomy: 20% postmenopausal	Laparotomy 6% (1.8 to 17.5%) Laparoscopy 0% (0 to 10.6%)	No malignancies Small sample size
Non-randomized comparisons				
Hidlebaugh et al., 1997[227]	405	199 laparoscopy 206 laparotomy 20.2% postmenopausal	Laparotomy 27.2% (21.8 to 34.0%) Laparoscopy 2.5% (1.0 to 6.0%)	Selection criteria for laparoscopy not defined Potential other risk factors for complications not described
Yuen et al., 1997[239]	110	Laparotomy: 6% postmenopausal Laparoscopy: 3.8% postmenopausal	Laparotomy 28% (18.5 to 43.1%) Laparoscopy: 9.6% (4.2 to 21.8%)	Difference between complication rates attributable to higher number of postoperative complications in laparotomy group
Carley et al., 2002[221]	106	44 laparotomy 62 laparoscopy Menopausal status not reported	Laparotomy 4.6% (0.7 to 16.7%) Laparoscopy 0% (0 to 8.6%)	
Chapron et al., 1997[222]	186	121 laparoscopy, 65 laparotomy 43% postmenopausal	Laparotomy: 15.4% (8.9 to 27.0%) Laparoscopy: 8.3% (4.6 to 15.0%)	Patients with high suspicion of malignancy went directly to laparotomy Results not analyzed by "intention to treat"—19 of laparotomy patients started as laparoscopy 13.6% of laparoscopies converted to laparotomy
Laparoscopy only				
Childers et al., 1996[32]	138	Not described in detail; age range 9-91	10.1% (6.2 to 16.7%)	Length of followup not given for benign cases

Study	N	Patient population	Complication rate (95% CI)	Notes
				Gynecologic oncology service Results not stratified by age or menopausal status 8.0% conversion to laparotomy
Canis et al., 1994[219]	757	11.4% postmenopausal	1.1% (0.53 to 2.1%)	Mean followup 42 months (range 3-153 months)
Dottino et al., 1999[37]	160	53% postmenopausal	7.5% (4.3 to 12.9%)	Gynecologic oncology service
Marana et al., 2004[230]	620	All less than 40 years old	0.9% (0.4 to 2.0%)	Mean followup 30 months Single surgeon
Parker et al., 1994[41]	61	100% postmenopausal	3.3% (0.4 to 12.3%)	Masses "presumptively benign" based on imaging, exam, clinical history 4.9% conversion
Sadik et al., 1999[232]	220	3.2% postmenopausal	0.9% (0.06 to 3.5%)	Malignant masses "excluded from study"
Chi et al., 2004[223]	146	Menopausal status not reported; median age 54	Mortality 2.5% (0.5 to 6.3%) Morbidity 22.1% (15.1 to 32.7%)	Clinical history not described—not clear if other conditions besides adnexal mass included
Havrilesky et al., 2003[226]	396	37.2% postmenopausal	Laparoscopy 8.3% (6.0 to 11.6%)	Risk of complication increased with concurrent hysterectomy
Lok et al., 2000[228]	513	5.5% postmenopausal	Laparoscopy 13.3% (10.6 to 16.6%)	No malignancies 75.% symptomatic
Mann and Reich, 1992[229]	44	100% postmenopausal	Laparoscopy 4.6% (0.7 to 16.7%)	1/44 had cancer
Parker and Proietto, 1997[231]	86	Menopausal status not reported	Laparoscopy 22.1% (15.1 to 32.7%)	1/86 had cancer
Serur et al., 2001[233]	100	49% postmenopausal	Laparoscopy 10% (5.6 to 19.0%)	-
Shalev et al., 1994[234]	55	100% postmenopausal	Laparoscopy 10.9% (5.2 to 22.9%)	-
Tarik and Fehmi, 2004[237]	1478	Menopausal status not reported (but mean age 30)	Laparoscopy: Diagnostic procedures 1.8% (0.8 to 3.8%) Minor procedures: 1.4% (0.8 to 2.3%)	Proportion with preoperative diagnosis of adnexal mass not reported
Van Herendael et al., 1995[238]	121	Menopausal status not reported	Laparoscopy: 1.7% (0.1 to 6.4%)	-

Abbreviation: CI = confidence interval

Nationwide Inpatient Sample

Table 25 shows the estimated numbers of discharges in the United States in 2000-2001 under each diagnostic class and procedure (standard errors not shown for simplicity). The results illustrate the difficulty in using discharge data to attempt to estimate morbidity and mortality rates for surgical procedures. Both morbidity and mortality are highest for cancer diagnoses, but

there is no way to determine the extent to which the underlying disease process contributed to either complications or death; for example, "exploratory laparotomy" or "diagnostic laparoscopy" in many ovarian cancer patients likely represents a "second-look" procedure done to determine response to chemotherapy. Outcomes of these procedures are not relevant to estimating the risks of a primary diagnostic procedure. The laparoscopies that are included in the NIS are likely not representative of all laparoscopies for adnexal masses; since the NIS does not capture surgeries performed at ambulatory surgery centers, the cases within the NIS may represent those for which surgeons had a higher index of suspicion of malignancy, or anticipated higher technical difficulty. Another major limitation is the inability to distinguish between the initial indication for surgery and the final diagnosis. Finally, in order to try to eliminate confounding by additional procedures, we excluded cases in which hysterectomy was performed – however, because hysterectomy is part of the standard initial surgical treatment of ovarian cancer, many cases of initial management are excluded.

Table 25. Estimated U.S. discharges for exploratory laparotomy and diagnostic laparoscopy with discharge diagnoses consistent with adnexal mass, with mortality and complication rates

	Number of discharges	Died	Mortality rate	Complica-tions	Complica-tion rate
OVARIAN CANCER	**118,042**	**7099**	**6.0%**	**515**	**0.4%**
Laparoscopy (no ovarian procedures)	222	5	2.3%	0	0.0%
Laparoscopy plus conservative ovarian procedure	27	0	0.0%	0	0.0%
Laparoscopy plus oophorectomy	16	0	0.0%	0	0.0%
Laparotomy (no ovarian procedure)	566	11	1.9%	5	0.9%
Laparotomy plus conservative ovarian procedure	68	0	0.0%	0	0.0%
Laparotomy plus oophorectomy	36	0	0.0%	0	0.0%
OTHER ADNEXAL CANCER	**780**	**15**	**1.9%**	**5**	**0.6%**
Laparoscopy (no ovarian procedures)	0	0	0.0%	0	0.0%
Laparoscopy plus conservative ovarian procedure	0	0	0.0%	0	0.0%
Laparoscopy plus oophorectomy	0	0	0.0%	0	
Laparotomy (no ovarian procedure)	15	15	100.0%	0	0.0%
Laparotomy plus conservative ovarian procedure	0	0	0%	0	
Laparotomy plus oophorectomy	0	0	0%	0	
BENIGN OVARIAN NEOPLASM	**145,024**	**255**	**0.2%**	**964**	**0.7%**
Laparoscopy (no ovarian procedures)	1,560	5	0.3%	35	2.2%
Laparoscopy plus conservative ovarian procedure	75	0	0.0%	0	0.0%
Laparoscopy plus oophorectomy	24	0	0.0%	0	0.0%
Laparotomy (no ovarian procedure)	700	4	0.6%	16	2.3%
Laparotomy plus conservative ovarian procedure	72	0	0.0%	0	0.0%
Laparotomy plus oophorectomy	31	0	0.0%	0	0.0%
PELVIC MASS	**13,625**	**30**	**0.2%**	**60**	**0.4%**
Laparoscopy (no ovarian procedures)					
Laparoscopy plus conservative ovarian procedure	41				
Laparoscopy plus oophorectomy					
Laparotomy (no ovarian procedure)	35	5	14.3%		
Laparotomy plus conservative ovarian procedure					
Laparotomy plus oophorectomy					
OVARIAN CYSTS	**474,485**	**376**	**0.08%**	**3045**	**0.6%**

	Number of discharges	Died	Mortality rate	Complica-tions	Complica-tion rate
Laparoscopy (no ovarian procedures)	5,508		0.00%	65	1.2%
Laparoscopy plus conservative ovarian procedure	274		0.00%		0.0%
Laparoscopy plus oophorectomy	173		0.00%		0.0%
Laparotomy (no ovarian procedure)	1,429	79	5.53%	19	1.3%
Laparotomy plus conservative ovarian procedure	99		0.00%		0.0%
Laparotomy plus oophorectomy	86		0.00%		0.0%
PARA-OVARIAN CYST	**21,807**	**5**	**0.0%**	**92**	**0.4%**
Laparoscopy (no ovarian procedures)	271		0.0%		0.0%
Laparoscopy plus conservative ovarian procedure	24	0	0.0%	0	0.0%
Laparoscopy plus oophorectomy	9	0	0.0%	0	0.0%
Laparotomy (no ovarian procedure)	61	10	16.4%	0	0.0%
Laparotomy plus conservative ovarian procedure	5	0	0.0%	0	0.0%
Laparotomy plus oophorectomy	5		0.0%		0.0%
PELVIC INFLAMMATORY DISEASE	**430,027**	**439**	**0.1%**	**4793**	**1.1%**
Laparoscopy (no ovarian procedures)	7,184	4	0.1%	150	2.1%
Laparoscopy plus conservative ovarian procedure	445	0	0.0%	9	2.0%
Laparoscopy plus oophorectomy	159	0	0.0%	5	3.1%
Laparotomy (no ovarian procedure)	2,129	10	0.5%	53	2.5%
Laparotomy plus conservative ovarian procedure	160	0	0.0%	0	0.0%
Laparotomy plus oophorectomy	45	0	0.0%	0	0.0%
NORMAL PELVIS	**108.8**	**0**	**0**	**0**	**0**

Other Outcomes

We identified two studies that reported on the sensitivity and specificity of intraoperative frozen section done to determine pathologic diagnosis.[220,236] They reported similar findings. Both studies defined low malignant potential tumors as cancer. Canis et al.[220] reported a sensitivity of 92.2 percent and a specificity of 92.2 percent in 141 women (29.8 percent postmenopausal, 35 percent with cancer or low malignant potential tumors). Tangjitgomol et al.[236] estimated similar values, with a reported sensitivity of 91.3 percent and specificity of 93.3 percent in 212 women (menopausal status not reported, cancer prevalence 77 percent). Defining low malignant potential cancers as benign decreased sensitivity in both cases.

We identified only one article that addressed the potential impact of surgical management of benign cysts on fertility. Somigliana et al.[235] followed 32 women who received ovarian stimulation after removal of an endometriotic cyst. The mean number of follicles observed in the ovary where the cyst had been removed (2.0 ± 1.5) was significantly lower than in the contralateral ovary (4.2 ± 2.5), suggesting that the surgical procedure may have led to decreased ovarian reserve. An alternative explanation is that the cyst itself had an adverse effect on ovarian reserve.

Discussion

Ideally, reports of adverse outcomes of diagnostic surgery for adnexal masses would be divided into four separate categories, based on preoperative symptoms and postoperative findings: (1) women with symptomatic masses which ultimately proved malignant; (2) women with symptomatic masses which ultimately proved benign; (3) women with asymptomatic masses which ultimately proved malignant; and (4) women with asymptomatic masses which ultimately proved benign. For the first three groups, the operative procedure could be considered appropriate even in the event of morbidity, since there is some benefit (primary surgical therapy for malignancy, or management of symptomatic nonmalignant adnexal pathology) to be gained from surgical diagnosis and treatment. For women with asymptomatic benign masses, there are theoretical benefits for detecting some benign masses, including (1) prevention of subsequent malignant transformation, (2) avoidance of rupture which, for certain benign masses (endometrioma and mature teratoma) could cause acute symptoms, (3) easier surgical management, with fewer complications, compared to management of a larger symptomatic mass, (4) avoidance of torsion (twisting of the adnexa) and emergent surgical management and (5) avoidance of effects on fertility, either from the underlying condition itself or from more extensive surgery for a larger mass. However, we did not identify any evidence for these benefits; the probabilities of these potential benefits also would differ widely depending on the underlying pathology and natural history of a particular mass, the patient's age and reproductive status, and other comorbidities.

Unfortunately, neither the literature nor available discharge data allow estimates of the probabilities of outcomes based on initial presentation. In the case of the literature, this is because of a lack of reporting of the clinical path by which patients come to undergo surgery. In the case of discharge data, it is because of the inherent limitations of the *International Classification of Diseases, Ninth Revision* (ICD-9) coding. Even if more recent data on ambulatory surgery were available, it would still be limited by coding.

Summary

Mortality for laparoscopic management of adnexal masses at experienced centers appears to be quite low, although the upper bound of this low rate is unclear.

Patient characteristics that determine risk of morbidity are unclear, although the need for more extensive procedures appears to increase the risk. Laparoscopy may have a lower morbidity rate than laparotomy, but this appears to be due, at least in part, to different patient selection criteria and surgical procedures performed.

Two small studies suggest that the false negative rate of intraoperative frozen section diagnosis is approximately eight percent, and the false positive rate is approximately five to seven percent. Whether either type of false result has a significant impact on outcome is unclear.

There is suggestive evidence that removal of a cyst in premenopausal women may affect ovarian reserve, potentially affecting fertility and/or age of menopause, but the underlying pathologic process may also play a role. More data are needed.

There are no data to allow estimation of the risks of a diagnostic procedure in the patient with an asymptomatic mass, or to assess the benefits of surgery in that patient compared to the risk of malignancy.

Question 7: Modeling Diagnostic Strategies

Question 7 is: What are the estimated trade-offs resulting from various strategies for evaluation of the adnexal mass?

Approach

A formal decision analytic approach is often quite helpful for synthesizing evidence coming from a range of sources, of varying quality, and of varying degrees of precision in estimates. Such models are also helpful in identifying which parameters are most important, in order to prioritize future research. Ideally, the underlying natural history of the disease in question can be modeled, with the impact of subsequent clinical interventions estimated based on test characteristics, effectiveness and morbidity from treatment, patient preferences, etc. In addition, the effect of varying both the incidence and natural history of ovarian cancer based on risk factors such as genetic predisposition can also be taken into account if adequate data are available. For example, such models have been quite helpful in exploring the impact of various interventions for cervical cancer prevention.[240] In addition, data from currently ongoing trials of ovarian cancer screening will also provide valuable data on natural history. [241]

Because of the methodological limitations of the literature on management of adnexal masses cited in the previous sections, a formal decision analysis does not seem appropriate at this time. In order to illustrate some of the key areas for future research, we did a simple estimate of the expected outcomes of several strategies for evaluation of the adnexal mass based on the findings of this review. Because models will ultimately need to incorporate the natural history of ovarian cancer, either to evaluate screening or to estimate the consequences of false negative diagnoses, we also performed a literature review of existing models of the natural history of ovarian cancer and the impact of screening or testing and developed an alternative model.

Predicting Outcomes of Management Strategies

As an example, we can consider one clinical scenario: an asymptomatic postmenopausal woman undergoing a routine bimanual pelvic examination. If the bimanual examination is abnormal, she can undergo a variety of additional tests. We compared several strategies: (1) performing CA-125 only, then operating on women with values greater or equal to 35;(2) performing an ultrasound with Doppler velocimetry (the strategy with highest sensitivity and specificity in our review) and operating on women with positive results both morphologically and with Doppler; (3) performing CA-125, then performing ultrasound with Doppler on women with elevated CA-125 and operating on women with positive ultrasounds; (4) performing ultrasound with Doppler first, then performing CA-125 on women with positive ultrasound results, and operating on women with elevated CA-125, and (5) performing both ultrasound and CA-125 and combining these results with menopausal status to use the RMI (discussed in detail under Question 4); women with RMI scores above the threshold undergo surgery. Strategies 3 and 4 are examples of serial testing, Strategy 5 an example of parallel testing. Table 26 provides estimates for key parameters based on the previous chapters; estimates for test characteristics are taken from the point estimates of the pooled random-effects models.

Table 26. Estimates for key model parameters

Parameter	Value
Prevalence of adnexal masses in postmenopausal women (Question 1)	Malignant: 0.1% Benign: 1.0%
Sensitivity of the pelvic examination to detect adnexal masses (Question 2)	0.45
Specificity of the pelvic examination to detect adnexal masses (Question 2)	0.90
Sensitivity of combined morphology and Doppler (Question 3)	0.86
Specificity of combined morphology and Doppler (Question 3) Note: We assumed that the specificity of ultrasound for determining the absence of pelvic mass was 100%.	0.91
Sensitivity of CA-125 in postmenopausal women (Question 3)	0.80
Specificity of CA-125 in postmenopausal women (Question 3)	0.87
Sensitivity of RMI (Question 4)	0.74
Specificity of RMI (Question 4)	0.91

Abbreviation: RMI = Risk of Malignancy Index

At the initial pelvic examination, the probability of detecting a mass equals:

> Probability of true positive test + Probability of true negative test, or
> (Prevalence of mass * Test sensitivity) + (1-Prevalence of mass)*(1-Test Specificity)

Similarly, the probability of a negative test equals:

> Probability of true negative + Probability of false negative, or
> (1-Prevalence)*Test Specificity + Prevalence*(1-Sensitivity)

At the time of ultrasound, the "prevalence" of disease is equal to the positive predictive value of the preceding test, the ultrasound, or:

> Probability of true positive pelvic/(Probability of true positive pelvic + Probability of false negative pelvic)

Similar calculations were made for each test or combination of tests.

Table 27 shows the predicted outcomes (in terms of detected and missed cancers) of testing with either ultrasound morphology with Doppler velocimetry or CA-125 alone in a hypothetical cohort of 100,000 postmenopausal women.

Table 27. Predicted outcomes of ultrasound plus Doppler or CA-125 testing to determine surgical management in a hypothetical cohort of 100,000 postmenopausal women*

	Underlying pathology						
	Cancer	Benign mass	Normal	Total	Prevalence of malignancy among test positives	Proportion of all tests positive	Missed cancers
Baseline cases	100	1,000	98,900	100,000	0.1%		
Pelvic exam							
Positive	45	450	9,890	10,385			
Negative	55	550	89,010	89615	0.4%	10.4%	55
STRATEGY: CA-125 only							
CA-125							
Positive	36	59	1,286	1,380			
Negative	9	392	8,604	9,005	2.6%	15.3%	9
Surgery							
Positive	36			36			
Negative		59	1286	1,345		2.6%	
STRATEGY: Morphology/ Doppler only							
Morphology/ Doppler							
Positive	39	41	0	80			
Negative	6	410	9,890	10,306	49.8%	0.8%	6
Surgery							
Positive	39	0	0	39			
Negative	0	41	0	41		49.8%	

* Some numbers may not add up correctly because of rounding.
Abbreviation: CA-125 = cancer antigen 125

Table 28 shows the predicted outcomes of the serial and parallel testing strategies.

Table 28. Predicted outcomes of serial testing or parallel testing with ultrasound plus Doppler or CA-125 testing to determine surgical management in a hypothetical cohort of 100,000 postmenopausal women*

	Underlying pathology						
	Cancer	Benign mass	Normal	Total	Prevalence of malignancy among positive tests	Proportion of all tests positive	Missed cancers
Baseline cases	100	1,000	98,900	100,000	0.1%		
Pelvic exam							
Positive	45	450	9,890	10,385			
Negative	55	550	89,010	89,615	0.4%	10.4%	55
STRATEGY: CA-125, followed by morphology/ Doppler							
CA-125							
Positive	36	59	1,286	1,380			
Negative	9	392	8,604	9,005	2.6%	13.2%	9
Morphology/ Doppler							
Positive	32	5	0	37			
Negative	4	53	1,286	1,343	86.5%	2.7%	4
Surgery							
Positive	32	0	0	32			
Negative	0	5	0	5			
STRATEGY: Morphology/ Doppler followed by CA-125							
Morphology/ Doppler							
Positive	40	41	0	81			
Negative	5	410	9,890	10,305	49.4%	0.8%	5
CA-125							
Positive	32	5	0	37			
Negative	8	35	0	43	86.5%	45.7%	8
Surgery							
Positive	32			32			
Negative	0	5	0	5		86.5%	
STRATEGY: RMI (morphology + CA-125 + menopausal status)							
RMI							
Positive	33	41	0	74			
Negative	12	410	9,890	10,312	44.6%	13.2%	9
Surgery							
Positive	33	0	0	33			
Negative	0	41	0	41		44.6%	

* Some numbers may not add up correctly because of rounding.
Abbreviations: CA-125 = cancer antigen 125; RMI = Risk of Malignancy Index

Table 29 summarizes the outcomes of the five strategies in terms of total number of tests, total number of missed cancers, and total number of surgeries.

Table 29. Estimated numbers of tests, missed cancers, and surgeries for each strategy

	Strategies				
	Single tests		Serial tests		Parallel tests
	CA-125	Ultra-sound*	CA-125 followed by ultrasound	Ultrasound followed by CA-125	Risk of Malignancy Index
Total tests	10,385	10,385	11,765	10,466	20,770
Total missed cancers	9	9	13	13	9
Total surgeries	1,380	80	37	37	74

Abbreviation: CA-125 = cancer antigen 125

Table 30 illustrates the effect of increasing the prevalence of cancer (for example, in symptomatic women with a known mass) from 0.1 percent to 10 percent. The size of the cohort here is 1,100 women with masses (the same as in the screening cohort).

Table 30. Estimated numbers of tests, missed cancers, and surgeries for each strategy in 1,100 women with known adnexal mass and underlying prevalence of ovarian cancer 10%

	Strategies				
	Single tests		Serial tests		Parallel tests
	CA-125	Ultra-sound*	CA-125 followed by ultrasound	Ultrasound followed by CA-125	Risk of Malignancy Index
Total tests	1,100	1,100	1,317	1,287	2,200
Total missed cancers	20	15	32	32	26
Total surgeries	197	184	90	90	155

Abbreviation: CA-125 = cancer antigen 125

This simple "model" illustrates several key points:

- The prevalence of malignancy increases as additional diagnostic tests are performed. This is certainly clinically appropriate and reflects the effects of sequential testing strategies. However, specificity and, to some extent, sensitivity for many of the tests reviewed appear to vary with underlying disease prevalence. Thus, estimates for test characteristics calculated at one point in the clinical pathway may not be appropriate for other points.
- Despite a poor sensitivity of 45 percent, the negative predictive value of a negative pelvic examination for malignancy is quite high (99.94 percent). The reassurance provided by a "normal" exam reflects the epidemiology of the underlying disease, rather than the intrinsic value of the test in discriminating benign from malignant. This reflects the low prevalence of ovarian cancer in the population. Conversely, the positive predictive value is only 0.4 percent, despite a specificity of 92 percent.
- In order to judge the trade-offs between detection of masses that ultimately prove malignant compared with the risks of diagnostic surgery, we would need better estimates of morbidity and mortality within different diagnostic categories – as noted previously, these do not exist.
- The most "efficient" strategy in terms of number of tests and surgeries is serial testing with ultrasound followed by CA-125; however, this results in four missed cancers compared with parallel testing using the RMI. However, parallel testing doubles the number of tests to be

performed. A formal cost-effectiveness analysis requires significantly more data on test characteristics and ovarian cancer natural history, as well as the morbidity of surgical management.

- Modeling parallel testing beyond the data in scoring systems is difficult. Besides requiring specific assumptions about how results that were positive for one test but negative for another would be managed, one would also need to know if the sensitivity and specificity of each test were independent or correlated in some way. For example, it seems likely that the sensitivity of both ultrasound and CA-125 would be greater for larger masses than for smaller masses.
- In scenarios where the likelihood of ovarian cancer is higher, the negative predictive value of any diagnostic strategy will decrease (more missed cancers), and the positive predictive value will increase (the proportion of surgical cases where cancer is found will be higher). This is seen clearly by comparing Tables 29 and 30. The number of women with adnexal masses is the same, but the number of missed cancers is substantially higher with each strategy.
- In addition, for any screening modality, there needs to be evidence that early detection reduces disease-specific morbidity and mortality. In addition, in order to judge the impact of false negative results, data on the natural history of ovarian cancer are also needed. Since data from large trials are still pending, one way to examine the potential impact of different testing strategies for both initial screening and subsequent testing is through the development of simulation models.

We next review published models of the natural history of ovarian cancer.

Models of Ovarian Cancer: Literature Review

Four articles were identified from the literature review that used modeling to determine the effectiveness and cost-effectiveness of different screening strategies for the detection and treatment of ovarian cancer. These are described in Evidence Table 7 (Appendix D*). Studies were included if they were directly relevant to Question 7,[242-244] or provided natural history information that could be used in the construction of a model.[245]

Schapira et al.[242] conducted a decision analysis comparing a one-time screen using transvaginal sonography and CA-125 either alone or in combination to determine life-expectancy gains in a cohort of 40-year-old women in the United States. In the model women could either be screened or unscreened. Probabilities were derived from the literature for the following: prevalence of disease in 40-year-old women, percentage of early stage disease, clinical detection of disease, sensitivity of the screening test for detection of early stage disease, specificity of the screening test, and the mortality rate associated with diagnostic laparotomy. Life expectancy was calculated for women who had no disease, early stage disease, and late stage disease. Table 31 summarizes key input parameters and ranges.

Assumptions in the model were that survival time for clinically and screen-detected early stage disease is the same; morbidity and mortality rates associated with diagnostic laparotomy are the same for people with and without the disease; and there is no benefit gained from identifying benign disease.

* Appendixes cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/ adnexal/adnexal.pdf.

The results of the analysis suggested that use of the combined strategy would result in a gain in life expectancy (compared to no screening) of one third of a day of life. No screening was preferred if the postoperative mortality rate exceeded 7.32 percent or the specificity of the test was less than 98.53 percent. An additional analysis, examining the use of testing for women aged 65+ suggested that the combined strategy would result in an average gain in life expectancy of approximately 3/4 of a day of life.

Table 31. Key input parameters and ranges for the Schapira model[242]

Parameter	Value	Range	Source
Prevalence of ovarian cancer	28.6/100,000	20 to 200/100,000	NCI monograph No. 41; 1975
Percentage of prevalent cases in early stage	50%	20 to 80%	Assumed
Percentage of early stage disease diagnosed clinically	25%	20 to 80%	ACS Cancer Statistics 1990
Sensitivity of CA-125 and TVUS (combined) for early stage disease	45%	20 to 80%	Literature review
Sensitivity of CA-125 and TVUS (combined) for late stage disease	81%	50 to 100%	Literature review
Specificity of CA-125 and TVUS	99.95%	96 to 100%	Literature review
Probability of post-laparotomy death	0.23%	0 to 10%	National Halothane Study JAMA 1966

Abbreviations: ACS = American Cancer Society; NCI = National Cancer Institute; TVUS = transvaginal ultrasound

Skates and Singer[244] developed a stochastic model to evaluate screening with CA-125. Key assumptions in this model included:

- Stepwise progression from Stage I through Stage II through Stage III through Stage IV;
- Log-normal distributions of progression rates;
- Stage at clinical detection independent of duration of disease;
- The coefficient of variation in stage length is constant across all stages;
- Estimates for the duration of each stage were provided by two gynecologic oncologists.

In the base case, the model predicted that screening would save 3.4 years of life per detected case; of note, estimates for the gains in life expectancy for the entire population undergoing screening were not provided.

Urban et al.[243] examined the cost-effectiveness of screening using CA-125 and TVUS alone or in combination in a cohort of 1 million 50-year-old women using a stochastic simulation model, building on the model of Skates and Singer (Table 32). Screening and case ascertainment was assumed to occur over a 3-year period; women were assumed to be followed until age 80 or death.

Table 32. Key assumptions and data sources used to derive values for parameters in the Urban model[243]

Parameter	Estimate	Source
Stage of ovarian cancer	FIGO	
Relative stage lengths (relative to Stage 1)	0.5, 1.333, 0.333	Skates et al.[244] FIGO stages III and IV assumed to comprise SEER stage 3
Geometric mean stage length in months	9; 4.5, 12 and 3 months	
Probability of disease during testing period	0.0121	Not stated
Probability of age at clinical detection	Age 50-54 – 0.153 Age 55-59 – 0. 184 Age 60-64 – 0.202 Age 65-69 - 0.179 Age 70-74 – 0.150 Age 75-80 – 0.132	SEER
Probability of stage at clinical detection	Stage 1 - 0.223 Stage 2 - 0.153 Stage 3/ 4 - 0.624	SEER
Point in stage at clinical detection	0.5 of stage length	Assumed
Stage length distribution	Log normal (9, 4.5)	Assumed
TVUS sensitivity	100%	van Nagell, CA 1990 van Nagell, CA 1991
TVUS – false positive	1st screen 0.019; 2nd screen 0.010; 3rd screen 0.006	Campbell, Br J Obstet and Gynecol 1990
CA-125 level in cases	Refer to page 254 of article for formula	Skates et al.[244] Einhorn, Proc Am Soc Clin Oncol 1990
% of false negatives for CA-125	5%	Assumption
CA-125 specificity in women with false positive TVUS	0.85	Bast, Gyn Onc 1985 Woolas, JNCI, 1993
Return to normal life-expectancy post-diagnosis	15 years	Assumption
Probability of death in surgery among false-positive	0.001	Assumption

Abbreviations: FIGO = International Federation of Gynecology and Obstetrics; SEER = Surveillance, Epidemiology, and End Results; TVUS = transvaginal ultrasound

Six screening strategies using TVUS and CA-125 either alone or in combination: annual TVUS; annual CA-125, elevated (35U/ml used for referral to laparoscopy); annual CA-125, rising or elevated (rising defined as CA-125 level that has doubled since last screen); annual TVUS conditional on rising or elevated CA-125; 6-month TVUS condition on rising or elevated CA-125; 2-year TVUS conditional on rising or elevated CA-125. Of these, the strategy of

annual TVUS conditional on rising or elevated CA-125 was identified as efficient, meaning it saved an equivalent if not higher amount of life at lower costs compared to other strategies. The model was especially sensitive to assumptions about the duration of Stage I disease.

Discussion

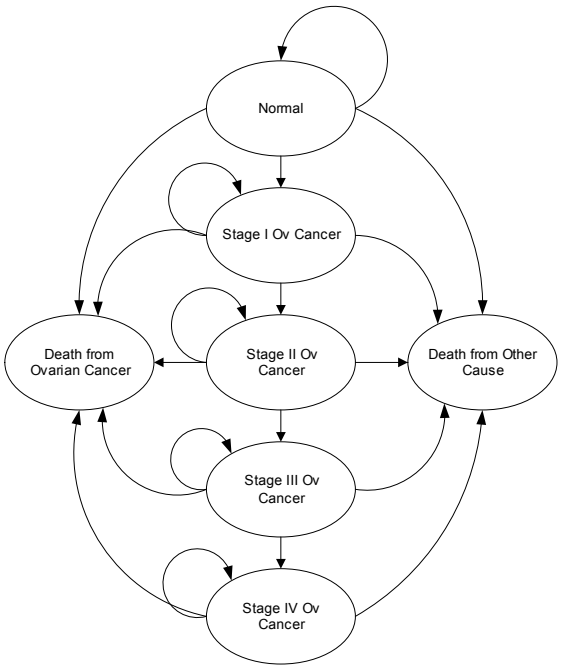
Secondary prevention of cancer mortality through screening has been remarkably effective in the case of cervical cancer. Mammography has also reduced mortality from breast cancer, although there remains some controversy. To date, although survival in early stage ovarian cancer is considerably higher than survival in later stage cancers, trials of screening have not yet demonstrated reduction in disease-specific mortality. Although the relative lack of effectiveness of ovarian cancer screening to date may reflect the lack of an appropriate test, differences in the biology and natural history of the different cancers may also result in some of the differences.

As outlined in a recent review,[246] the most critical criteria for an effective screening strategy for ovarian cancer is that there is a time of sufficient duration during the development of ovarian cancer when cancer is detectable but in a stage when treatment effectiveness is high. As shown in the two most sophisticated models reviewed, estimates of the effectiveness of screening are highly dependent on assumptions about the duration of Stage I cancer. The basis for the estimates used in both models was the opinion of two clinicians; the methods used to derive these estimates were not described.

Cervical cancer is, in the majority of cases, a squamous carcinoma, which spreads primarily through direct extension and secondarily through lymphatic invasion. The most common type of ovarian cancer, on the other hand, is typically an adenocarcinoma, which spreads by dissemination of tumor cells throughout the peritoneal cavity.

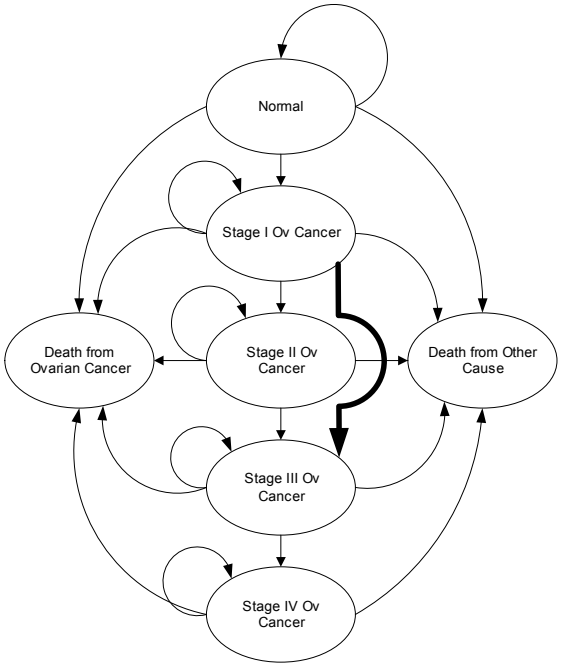
One assumption commonly made in the models of ovarian cancer we identified is that ovarian cancer staging represents the natural history. Figure 24 illustrates a simplified schematic model used in all three of the reviewed papers. Patients can develop ovarian cancer, die of other causes, or remain healthy. Those who develop ovarian cancer can present with symptoms or through testing to become an incident case, or remain undetected, and can either remain within the same stage or progress to the next.

Figure 24. Schematic of Markov or stochastic model of ovarian cancer natural history



Although this stepwise progression through stages is the case for cervical cancer, there is no evidence to suggest that tumors limited to the ovary (Stage I) must necessarily spread first to adjacent pelvic organs (Stage II) prior to spread throughout the peritoneal cavity (Stage III). Although staging systems represent the extent of disease, they are developed to help with prognosis, and to allow comparison of treatment effectiveness – there is no explicit assumption that each stage necessarily must be preceded by the next lowest one. Figure 25 depicts an alternative model, which allows some Stage I cancers to progress directly to Stage III:

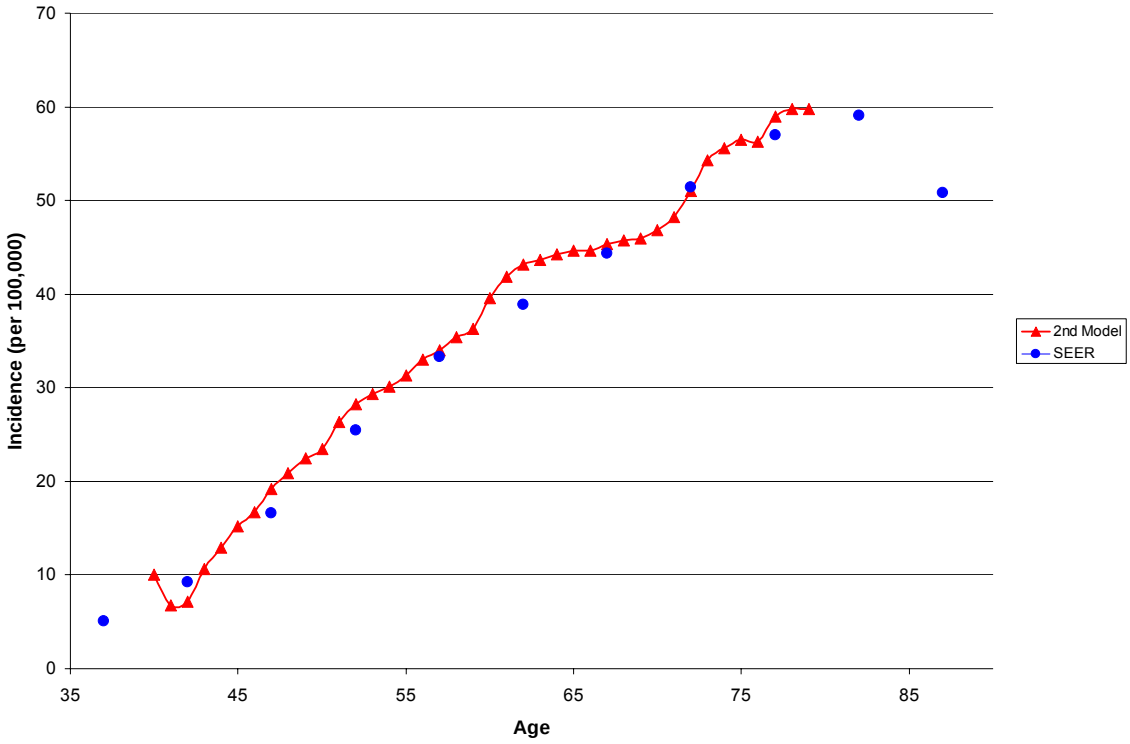
Figure 25. Alternative model of ovarian cancer history



Using the Markov model described in Chapter 2, we performed sensitivity analyses on progression rates and type of progression to determine if this second "model" of progression could result in similar stage distributions to observed data.

Figure 26 compares the predicted incidence of ovarian cancer derived from the model with incidence rates reported in the Surveillance, Epidemiology, and End Results (SEER) data set, under the assumption that there was a stepwise progression from Stage I through Stage IV:

Figure 26. Model predictions of ovarian cancer incidence (black triangles) compared to SEER incidence rates (closed circles)



We then allowed a proportion of Stage I cancers to proceed directly to Stage III and calibrated underlying progression rates. Table 33 compares the model input parameters and resulting stage distribution of the two models.

Table 33. Inputs and outputs of ovarian cancer models

	Model 1 (Stage 1 must progress through Stage II)	Model 2 (some Stage I can progress directly through Stage III)	Stage distribution: FIGO (local data from Skates et al.[244])	Stage distribution: SEER (1995-2001)
Parameter estimate				
Annual probability of presenting with symptoms: Stage I	0.095	0.1		
Annual probability of presenting with symptoms: Stage II	0.095	0.15		
Annual probability of presenting with symptoms: Stage III	0.7	0.9		
Annual probability of presenting with symptoms: Stage IV	1	1		
Proportion of Stage I progressing directly to Stage III	0	0.25		
Model output: stage distribution				
FIGO:				
Stage I	19.1%	19.6%	25%	
Stage II	8.2%	9.3%	8%	
Stage III	54.2%	65.2%	52%	
Stage IV	18.6%	5.9%	15%	
SEER/WHO:				
Local	19.1%	19.6%	25%	19%
Regional	8.2%	9.3%	8%	7%
Distant and unstaged	72.8%	71.1%	67%	75%

Abbreviations: FIGO = International Federation of Gynecology and Obstetrics; SEER = Surveillance, Epidemiology, and End Results; WHO = World Health Organization

With relatively small changes in the probability of presenting with symptoms, a model that allows 25 percent of Stage I tumors to progress directly to Stage III results in stage distributions similar to observed data, and results in similar lifetime risk of ovarian cancer as the Urban model,[243] In a model with multiple input parameters, a huge number of combinations of parameters can result in similar outputs. Given that estimations of the duration of the different stages of ovarian cancer are based on little empirical data, and that there is no empirical data on the natural history of ovarian cancer, further exploration of the implications for screening, and the evaluation of masses detected through screening, is warranted.

Summary

The evidence is insufficient to develop a comprehensive model to estimate the relative benefits and risks of different management strategies for evaluating the adnexal mass.

Based on summary estimates of pooled sensitivity and specificity, management strategies that use imaging as the first step for evaluating an adnexal mass detected on examination (as opposed to CA-125) are more efficient, since they exclude false positive results from further examination. Serial testing with imaging followed by CA-125 results in the fewest number of

surgeries, but misses more cancers than parallel testing. Parallel testing greatly increases the number of tests required, but results in fewer missed cancers. Additional data are needed to evaluate cost-effectiveness.

Alternative assumptions about the natural history of ovarian cancer can result in modeled outcomes similar to those of published models; the implications of these assumptions should be explored further.

Chapter 4. Discussion

Limitations of the Report

There are several limitations to this evidence report:

- We did not review articles published in languages other than English because of a lack of resources for translation. It is possible that this led to the failure to include some relevant articles.
- For our review of prevalence studies (Question 1), we excluded studies performed outside the United States. Because the report was requested by the Centers for Disease Control and Prevention (CDC) to help with development of their policies and research agenda into ovarian cancer prevention strategies, we focused on U.S. populations and reasoned that the underlying prevalence of different conditions in women with adnexal masses could well differ in potentially important ways due to differences in racial/ethnic distribution and/or environmental exposures. As discussed in Chapter 2, this is supported by wide international variation in the incidence of cancer. Variations in screening, diagnosis, and surgical management could also lead to differences in the prevalence of various conditions among women with adnexal masses. It is possible that this reasoning was incorrect, and that some relevant articles were excluded. However, some non-U.S.-based articles were reviewed for other questions, and the majority shared the same biases as U.S.-based studies (i.e., most were done immediately preoperatively).
- There was considerable heterogeneity in design and patient populations among studies, and our use of a random-effects model to perform meta-analyses for some questions may have led to inaccurate estimates of pooled sensitivity and specificity. We also did not weight the results by anything other than sample size; it is possible that different results might have been obtained by weighting for study quality, for example.
- In our review of data from the Nationwide Inpatient Sample, we used only specific *International Classification of Diseases, Ninth Revision* (ICD-9) "E" class codes to identify complications. A more exhaustive strategy (e.g., identifying procedures not typically performed at the time of diagnostic surgery, identifying blood transfusions through procedure or charge codes, including patients with cancer who underwent hysterectomy) might have revealed more complications,[26] but would have required additional assumptions about the original indication for the surgery and the likely potential contribution of different aspects of the procedure to the complication (e.g., hysterectomy vs. oophorectomy).
- Our exploration of alternative models for the natural history of ovarian cancer did not directly compare estimated outcomes of screening strategies to other models. However, a comprehensive evaluation of screening for ovarian cancer was beyond the scope of this report. We are currently developing the model further to conduct these analyses.

Methodological Issues in the Literature

Description of the Patient Population

The main shortcoming of many of the papers reviewed was a failure to adequately describe the patient population, including the manner in which the adnexal mass was originally detected and subsequent evaluation. In Chapter 1, we described the importance of understanding the clinical presentation of the subjects in studies of management of adnexal masses. Because prevalence directly affects predictive values and may indirectly affect estimates of sensitivity and specificity, the probability that a patient is a true or false positive, or true or false negative, is dependent on the prevalence. In addition, the presence or absence of symptoms can affect the probability that a patient will undergo surgery if test findings indicate a benign mass, since surgery may still be the treatment of choice for the underlying condition. We were disappointed that the overwhelming majority of the studies we reviewed, relevant to all of the questions, did not adequately describe their population, so that the proportions of patients who presented with asymptomatic masses versus those with symptoms could be compared.

To be fair, there is an inherent feasibility issue in studies of diagnostic test accuracy for ovarian cancer – the ideal reference standard is histological confirmation, yet this confirmation requires surgery. Although this is a limitation of all cancer screening tests, the surgery required for a definitive diagnosis of ovarian cancer is more extensive than that for many cancers (for example, cervical, breast, and colon cancer can all be diagnosed without a requirement for general anesthesia). Especially with screening, or early in the diagnostic evaluation, the risks of surgery may be difficult to justify (especially since the low prevalence of malignancy makes the positive predictive value of tests early in the diagnostic evaluation quite low). From a research ethics perspective, it is certainly reasonable to limit diagnostic test studies to patients already scheduled for surgery. However, readers of these studies should recognize that the prevalence of malignancy will be substantially higher in preoperative patients than in patients at the time of the initial diagnosis of adnexal mass. Because test performance may be affected by prevalence, the outcomes (in terms of true and false test results) may be quite different in these two patient populations.

The same caveats hold for studies of the outcomes of surgery. Morbidity and mortality related to surgical diagnosis are influenced by the underlying diagnosis, as well as the extent of the disease (such as size of the mass, presence of adhesions from the disease process or prior unrelated surgery, or cancer stage). Interpreting surgical outcomes from studies that do not provide relevant clinical information is difficult; at the least, generalizablity is a major concern. Lack of relevant clinical information is a particular problem with administrative databases, which otherwise have the attraction of large sample size and better generalizability.[26]

An even more basic shortcoming was the failure to describe potential differences in study results stratified by age or menopausal status. Given the clear and widely recognized relationship between age and ovarian cancer risk, all studies in this area should present results in a way that allows separate estimation of outcome by age/menopausal status.

Sample Size

Few of the studies we reviewed included a priori sample size calculations. Use of confidence intervals for parameter estimates was uncommon. In studies of scoring systems, there were often too few cases of cancer for the number of variables included in the original models.

Blinding

Relatively few of the diagnostic studies reported whether those interpreting test results were blinded to either clinical presentation or ultimate diagnosis. This could clearly have an impact, particularly in studies of the bimanual pelvic examination; the finding that specificity decreased as prevalence increased suggests that the threshold for identifying a mass as cancer is lower if the clinical suspicion – based on other factors such as patient age, menopausal status, or history – is higher. Although this may be appropriate clinically, it results in biased estimates of test performance.

Observer Variability

Few studies addressed the potential impact of observer variability on the precision of test characteristics.

Natural History of Ovarian Cancer

As discussed in more detail in the section on Question 7, ovarian cancer has been implicitly assumed to progress through a series of stages in a way analogous to cervical cancer. Alternative models are biologically plausible, and mathematical models can be "fitted" to match reported data under a variety of scenarios. Since existing models already show that the effectiveness of screening is dependent on assumptions about the length of Stage I, further exploration of the impact of varying assumptions about natural history is warranted.

The most important parameter in these models, stage duration, is inherently unknowable; however, the source for the parameter estimate in the two most sophisticated models were "personal communications" with two gynecologic oncologists. At the least, more formal methods of eliciting expert opinion are probably warranted for future modeling studies.

Implications of Findings

Question 1

The prevalence of malignancy, even in postmenopausal women, is low – approximately 0.1 percent (1 in 10,000) in large screening studies in the United States. The potential for screening to reduce morbidity and mortality is currently being tested in at least three large trials; these trials should also provide valuable data on disease prevalence and the effectiveness of various followup strategies.

Question 2

Until the results of the large screening trials are available, many, if not most, women with asymptomatic adnexal masses will have had the mass detected as part of a routine health maintenance examination.

The bimanual pelvic examination appears to have a sensitivity of less than 60 percent, whether for detecting adnexal masses in general or for distinguishing benign from malignant masses. Based on the best pooled estimate of sensitivity (45 percent) and a prevalence of 0.1 percent, a normal risk, asymptomatic, postmenopausal woman with a normal pelvic examination has a 99.94 percent chance of not having cancer, even though over half of the cancers would be missed. This is due to the low prevalence of ovarian cancer, since, even without the test, her probability of not having cancer is 99.99 percent. Given these test characteristics, the value of the pelvic examination in reducing ovarian cancer morbidity and mortality appears to be extremely limited, at best. Although there may be some rationales for an annual bimanual examination (discussed in Chapter 5), ovarian cancer screening is not one of them.

Question 3

Of the various diagnostic imaging modalities, either a combination of ultrasound morphology and Doppler velocimetry, or magnetic resonance imaging (MRI), had the best combination of sensitivity and specificity for distinguishing benign from malignant disease. If confirmed by direct comparison, cost-effectiveness might be the most important determinant of which would be the optimal diagnostic procedure. Because the specificity of cancer antigen 125 (CA-125) is high in postmenopausal women, it is helpful in ruling in disease.

Question 4

Additional validation of scoring systems in new populations is required before widespread adaptation can be recommended.

Question 5

The most effective and efficient method for following patients who have been classified as having a benign mass is unclear, although unilocular cysts less than 10 cm appear to have a very low risk of malignancy.

Question 6

The risks of diagnostic laparoscopy or laparotomy, particularly in asymptomatic women who ultimately prove to have a benign lesion, are unclear. Overall morbidity appears to be low in reported series, but these are subject to numerous biases, particularly regarding selection for laparoscopy. Two small randomized trials suggest higher short-term morbidity with laparotomy compared to laparoscopy, but differences between the two groups raise the possibility of confounding.

Question 7

Based on our pooled estimates of sensitivity and specificity, serial testing of postmenopausal women with an adnexal mass detected by pelvic examination with either ultrasound morphology plus Doppler imaging, or MRI (which had similar sensitivities and specificities), followed by CA-125, resulted in the most efficient combination of number of tests, missed cancers, and surgeries. Parallel testing and using a scoring system such as the Risk of Malignancy Index resulted in fewer missed cancers than serial testing, but more overall tests and more surgeries. Additional data are needed to refine these estimates, to include the morbidities of the tests and surgeries, and to perform cost-effectiveness analyses. Either combined strategy is preferable to using imaging alone or CA-125 alone.

We cannot directly compare these results to the joint guidelines of the Society of Gynecologic Oncologists (SGO) and American College of Obstetricians and Gynecologists (ACOG) on which patient to refer to a gynecologic oncologist[247] because the data were not available to replicate their findings. However, our results are consistent with the guidelines, which recommend a CA-125 level above 35 for postmenopausal women, the presence of ascites, or evidence of adnexal or distant metastasis.

Alternative assumptions and parameter estimates can be used to generate predicted cancer incidences similar to those seen in published models of the natural history of ovarian cancer. In order to better estimate the potential impact of different strategies for ovarian cancer screening, and for managing masses detected through screening or presenting with symptoms, additional models that explore the implications for alternative natural history assumptions are needed. Data from ongoing screening trials may provide estimates of many of the currently unknown parameters.

Chapter 5. Future Research

This section outlines research priorities identified through the review, both in terms of fundamental gaps in knowledge and in addressing methodological issues of existing studies.

Minimal Data Reporting

Our ability to stratify results by relevant patient characteristics, or to compare the potential effect of patient characteristics on different results from different studies, was limited by the lack of information in most studies. We would suggest that future studies relevant to the diagnosis and management of adnexal masses provide data on, and present results stratified by, the following minimum characteristics:

- Patient age and/or menopausal status
- Patient body mass index
- Patient race and ethnicity
- Presence or absence of risk factors for ovarian cancer, particularly family history
- Means by which the adnexal mass was initially diagnosed—pelvic examination or imaging
- Reason for the initial examination which led to diagnosis of mass: symptoms referable to pelvic mass or ovarian cancer, examination for other symptoms, asymptomatic screening for ovarian cancer, or asymptomatic screening for other conditions

Prevalence of Different Types of Adnexal Masses

- Large scale screening trials will provide some data on the prevalence of different types of masses.
- Administrative data from surgical procedures may provide crude estimates, but some important information (like stage and grade of cancer, or histologic subtype) will likely be missing. In addition, relevant clinical data on presence or absence of symptoms and the diagnostic pathway leading to diagnosis will likely be missing. The best resource for obtaining the necessary data would likely be a large health maintenance organization (HMO) or third-party payer, which would allow comparison of inpatient and outpatient records, and followup of patients after diagnosis. Medicare data would provide similar information for women 65 and older.
- Separate reporting of the prevalence of different types of masses among women with and without symptoms would be helpful for clinical decisionmaking.

Diagnostic Testing

- Ideally, tests would be evaluated at the stage in the clinical pathway in which they are to be used.

- Since this means that many women who have a negative test will not undergo the reference standard, careful attention should be paid to development of alternative reference standards, including definitions of appropriate length of followup.
- More direct comparisons of alternative tests should be performed; existing studies are frequently underpowered to detect clinically meaningful differences, or to establish equivalence. Based on pooled analyses, either magnetic resonance imaging (MRI) or combined ultrasound evaluation of morphology and Doppler velocimetry have attractive sensitivity and specificity. Only two studies, with a total of 200 subjects, have directly compared these modalities in the same patient population.[91,100] In both of these studies, MRI was less sensitive but more specific than combined morphology/Doppler. More precise comparative estimates should be obtained.
- There is a paucity of studies on positron emission tomography (PET) compared to other imaging modalities. Given that the Centers for Medicaid and Medicare Services (CMS) is now reimbursing for PET scans done within the setting of a clinical trial, there is an excellent opportunity for high-quality studies which avoid the deficiencies outlined in this report.
- Although discriminating between benign and malignant lesions is the highest priority in most clinical situations, estimates of the sensitivity and specificity of various imaging modalities for specific nonmalignant lesions (endometriomas, mature teratomas, etc.) would be helpful for developing comprehensive management strategies, particularly in conjunction with good data on prevalence in premenopausal women. We identified multiple articles relevant to this question during our search, which were excluded because they were not relevant to the main study questions. Although many of the methodological issues identified here would be issues with these studies, a systematic review of this literature would have value.
- New tumor markers should continue to undergo evaluation as diagnostic tests as they are identified, using appropriate methodological standards.

Scoring Systems

- Validation studies in new populations are needed.
- Attention should be paid to adequate sample size.

Followup Studies

- Additional studies, with clear definitions for "benign" lesions and clear protocols for followup, with documentation of loss to followup, are needed. Because by definition these types of studies will not have histological confirmation of all test results, estimates of test performance from such studies may have some bias.

Adverse Outcomes of Surgery

- As with studies of prevalence, both currently published studies (mostly case series) and administrative data have significant deficiencies. Case series would be improved by clearer description of the clinical pathway by which patients ended up undergoing surgery, as well as

by providing relevant clinical data (such as body mass index, history of prior surgeries, and extent of disease).

- Data on outcomes from a variety of settings, including community settings, are needed.
- Again, as with studies of prevalence, data from sources able to provide both inpatient and outpatient data over time, such as HMOs, third-party payers, and Medicare, are likely to provide the best combination of sample size, generalizability, and clinical detail.

Sensitivity and Specificity of the Pelvic Examination

- The annual bimanual pelvic examination appears to have little, if any, benefit for reducing ovarian cancer morbidity and mortality in asymptomatic women. Given that many organizations now recommend less frequent cervical cancer screening in many women, that no screening test has ever been shown to reduce morbidity and mortality from endometrial cancer, and that other gynecological cancers are too rare to justify population-based screening, it would appear that annual bimanual pelvic exams do not have a substantial benefit in reducing mortality. Therefore, evidence on the benefits of the exam would be helpful for patients, clinicians, and policymakers. Possible research areas include:
  - Many clinicians argue that the annual exam provides a "cue" for women to interact with a clinician and receive other preventive services.
    - Would women be less likely to see a health professional on a regular basis if they would not get a pelvic examination?
    - If the exam does provide a "cue" for some women, what is its effectiveness and cost-effectiveness compared to alternative methods of improving adherence to periodic health maintenance schedules?
    - Are there some women who do not regularly see a health professional because of embarrassment/fear/discomfort regarding a pelvic exam who would be more likely to see one if they could be assured they would not get an exam?
  - Others have argued that, after long experience, women expect to receive a pelvic examination (and Pap test) on an annual basis and will continue to demand the examination, despite evidence that the test has little benefit, or does not need to be performed on an annual basis.
    - How have patients reacted to other changes or paradigm shifts in medicine? Can patient expectations be changed in the face of new evidence? Do patient responses differ between changes in which one intervention is replaced by another, versus changes in which an intervention is no longer performed at all?
  - Although the pelvic examination does not appear to have significant benefit as a screening test, does it have more value as a diagnostic test?
  - Assuming the pelvic examination does have value as a diagnostic test, is there a relationship between volume/experience and test accuracy, as suggested by two of the studies we reviewed? If so, can routine examinations in asymptomatic women be justified as a method for maintaining exam skills?
  - If there is a relationship between volume and accuracy, what are the implications for the performance of diagnostic bimanual examinations by generalists (e.g., internists, pediatricians, family practitioners, generalist nurse practitioners) versus specialists (e.g., obstetrician/gynecologists, nurse-midwives, etc)

Modeling the Outcomes of Different Screening Strategies

- Our modeling of the likely outcomes of different screening strategies was limited by the quantity and quality of data available for key parameters. Because this limited direct comparison of different testing strategies, we were not able to do a comprehensive comparison. The lack of data on patient characteristics, particularly symptom status, also prevented extensive analysis of the effects of different strategies in different clinical scenarios. Improving the evidence base for the other questions considered in the evidence report will make a substantial improvement in the ability to meaningfully model outcomes.
- Data on relevant patient preferences for different outcomes are needed.

- Data on relevant cost parameters are needed for cost-effectiveness analysis.

- Data on relative test reproducibility can help determine the effect of observer variability on effectiveness and cost-effectiveness.

Modeling the Natural History of Ovarian Cancer

- We identified only three models, one of which was an updated version of another. Having several groups working on simulation modeling, using different assumptions, software, model structure, etc., has proven quite helpful in the case of cervical cancer. Additional work should be strongly encouraged.
- In particular, models should explore alternative disease natural history parameters, and the implications for various strategies, including screening and primary prevention.

Chapter 6. Conclusions

Developing an effective and efficient algorithm for the evaluation of any condition requires good evidence on the prevalence of the condition at the first diagnostic encounter, and the sensitivity and specificity of the potential diagnostic tests to be used. With this information, one can estimate the outcomes, in terms of true and false positive and negative results, of each test. Various combinations of tests can be compared, and, ideally, the consequences of each test's results in terms of benefits, harms, and costs can be estimated.

In the setting of an adnexal mass, the primary issue is discriminating benign from malignant masses; ideally, all women with an underlying ovarian malignancy would receive appropriate surgical management (perfect sensitivity), and no woman with an asymptomatic benign mass would undergo unnecessary surgery (perfect specificity). The optimal strategy may well differ based on whether or not the patient presents with symptoms, both because the prevalence of disease is likely to be higher in the patient with symptoms (making the positive predictive value higher and the negative predictive value lower), and because surgical management may ultimately be appropriate for a symptomatic patient, and some asymptomatic patients, even if the mass is benign. Age and/or menopausal status are also important considerations, primarily because ovarian cancer is rare prior to age 50, but also because some of the risks of surgery may increase with age.

Unfortunately, the overwhelming majority of the literature we reviewed did not provide sufficient detail on these important patient characteristics to allow confident estimation of the outcomes of different diagnostic strategies, so that we are unable to conclude that any of the strategies achieve the aims of maximizing appropriate treatment and minimizing unnecessary surgery. Outside of studies that were explicitly designed to evaluate screening, few articles described whether patients were symptomatic or asymptomatic, or testing done prior to the diagnostic test being evaluated. Surprisingly few studies reported results separately for premenopausal and postmenopausal women. Future studies need to provide this information.

All of the diagnostic tests and scoring systems we evaluated exhibited a trade-off between sensitivity and specificity – studies of a given test that reported higher sensitivity had lower specificity, and vice versa. In pooled analysis, either the combination of ultrasound morphology and Doppler blood flow, or magnetic resonance imaging (MRI), had the best combination of sensitivity and specificity. Simple modeling of series and parallel tests suggests that, in postmenopausal women, imaging using ultrasound morphology and Doppler blood flow, or MRI, followed by CA-125, is both more sensitive (misses fewer cancers) and more specific (avoids more surgery) than either test alone. A strategy in which both tests were performed and used in a scoring system, the Risk Malignancy Index, prevented additional cancers but with twice as many tests and more surgeries. More data on key parameters are needed to determine if, in certain settings, alternative combinations of tests, performed in parallel or series, might have better outcomes or be more efficient.

Studies of surgical management suffered from the same limitations in terms of description of patient characteristics, making estimation of the risks of false positive diagnostic testing impossible. Similarly, administrative data that only includes discharge information do not provide important clinical information.

The bimanual pelvic examination has low sensitivity for both detection of adnexal masses and discriminating benign from malignant masses, raising doubts about its utility as a screening test in asymptomatic women.

Ultimately, evaluation of potential strategies for reducing morbidity and mortality from ovarian cancer may require use of simulation models, a technique that has proven helpful in evaluating prevention strategies for other cancers. Because the natural history of ovarian cancer is relatively unknown, testing of alternative models is critical. Although a few sophisticated models exist, development of additional models would be helpful, especially in the context of evaluating results from ongoing trials of screening. If any of these trials show a benefit from screening, then the need for better evidence on the diagnostic evaluation of adnexal masses will become even more critical.

References Cited in the Evidence Report

1. Jemal A, Murray T, Samuels A, et al. Cancer statistics, 2003. Ca: Cancer J Clin 2003;53(1):5-26.

2. Anonymous. SEER Stat Databases: Incidence - SEER Regs Public - Use, Nov 2004 Sub for Expanded Races (1992-2002) and Incidence - SEER 13 Regs excluding AK Public - Use Nov 2004 Sub for Hispanics (1992-2002), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2005, based on November 2004 submission. Available at: www.seer.cancer.gov. Accessed: May 25, 2005.

3. Russell P, Robboy S, Anderson M. Ovarian tumors: classification and clinical perspective. In: Robboy SJ, Anderson MC, and Russell P, editors. Pathology of the female reproductive tract. London: Churchill Livingstone; 2002. p. 527-38.

4. Eheman C, Bobo J, Lawson H, et al. Identifying public health opportunities to reduce the burden of ovarian cancer: workshop summary. Atlanta, Ga. Centers for Disease Control and Prevention, 2001. Available at: www.cdc.gov/cancer/ovarian/workshop.htm#conclusion. Accessed: May 25, 2005.

5. Eheman C, Brustrom J, Lawson H. The use of ultrasound in diagnosing ovarian cancer: can we improve on current practice? workshop summary. Atlanta, GA: Centers for Disease Control and Prevention, 2002. Accessed: http://www.cdc.gov/cancer/ovarian/workshop.htm. Accessed: May 25, 2005.

6. Russell P, Robboy S, Anderson M. The ovary: normal appearances and non-neoplastic conditions. In: Robboy SJ, Anderson MC, and Russell P, editors. Pathology of the female reproductive tract. London: Churchill Livingstone; 2002. p. 475-526.

7. Barber H, Graber E. The PMPO syndrome (postmenopausal palpable ovary syndrome). Obstet and Gynecol. 1971;38(6):921-3.

8. Munstedt K, von Georgi R, Misselwitz B, et al. Centralizing surgery for gynecologic oncology--a strategy assuring better quality treatment? [see comment]. Gynecol Oncol 2003;89(1):4-8.

9. Junor E, Hole D, McNulty L, et al. Specialist gynaecologists and survival outcome in ovarian cancer: a Scottish national study of 1866 patients. Br J Obstet Gynaecol 1999;106(11):1130-6.

10. Tingulstad S, Skeldestad FE , Hagen B. The effect of centralization of primary surgery on survival in ovarian cancer patients. Obstet Gynecol 2003;102(3):499-505.

11. American College of Obstetricians and Gynecologists. ACOG Committee Opinion: number 280, December 2002. The role of the generalist obstetrician-gynecologist in the early detection of ovarian cancer. Obstet Gynecol 2002;100(6):1413-6.

12. American Academy of Family Physicians. Summary of recommendations for periodic health examinations. Leawood (KS):American Academy of Family Physicians. 2004.

13. Geomini P, Bremer G, Kruitwagen R, et al. Diagnostic accuracy of frozen section diagnosis of the adnexal mass: a metaanalysis. Gynecol Oncol 2005;96(1):1-9.

14. American College of Obstetricians and Gynecologists. ACOG Practice Bulletin 45. Cervical cytology screening. August 2003. 2003.

15. Whiting P, Rutjes A, Reitsma JB, et al. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern. Med 2004;140(3):189-202.

16. Gohagan J, Prorok P, Hayes R, et al. The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization and status. Control Clin Trials 2000;21 (6Suppl):251S-72S.

17. Prorok P, Andriole G, Bresalier R, et al. Design of the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial. Control Clin Trials 2000;21 (6 Suppl):273S-309S.

18. Jacobs I. European randomized trial of ovarian cancer screening (protocol). London: Wolfson Institute of Preventive Medicine, Department of Environmental and Preventive Medicine. 1995.

19. Goff B, Mandel L, Melancon C, et al. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. [see comment]. JAMA 2004;291(22):2705-12.

20. Thompson ISI ReseachSoft. ProCite [computer program]. Berkeley, CA: Thompson ISI ResearchSoft.

21. Nanda K, McCrory D, Myers E, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. Ann Intern Med 2000;132(10):810-9.

22. Lau J. Meta-Stat 0.6. Shareware program for performing meta-analyses of diagnostic tests. Available at: www.medepi.net/meta/software/MetaTest/Mtreadme.txt. Accessed August 26, 2005.

23. DerSimonian R , Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7(3):177-88.

24. Moses LE, Shapiro D. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993;12:1293-316.

25. Agency for Healthcare Research and Quality. Overview of the HCUP Nationwide Inpatient Sample. Rockville, MD. 2003.

26. Myers E, Steege J. Risk adjustment for complications of hysterectomy:utility of routinely collected administrative data. Am J Obstet Gynecol 1999;181:567-75.

27. Ries L, Eisner M, Kosary C, et al. SEER Cancer Statistics Review, 1973-1999. National Cancer Institute: Bethesda, MD. 2002.

28. Anderson R, DeTurk P. National vital statistics reports: United States life tables, 1999. National Center for Health Statistics: Hyattsville, MD. 2002.

29. Jones H. Ovarian Cancer: The Clinical Problem in Early Detection of Ovarian Carcinoma with Transvaginal Sonography: Potentials and Limitations, H.W. Jones, editor. Raven Press, Ltd.: New York. 1993;3-9.

30. TreeAge Software. DATA 4.0 [computer program]. Williamstown, MA: TreeAge Software, Inc.

31. Chalas E, Welshinger M, Engellener W, et al. The clinical significance of thrombocytosis in women presenting with a pelvic mass. Am J Obstet Gynecol 1992;166(3):974-7.

32. Childers JM, Nasseri A, Surwit EA. Laparoscopic management of suspicious adnexal masses. Am J Obstet Gynecol 1996;175(6):1451-7; discussion 1457-9.

33. Cohen LS, Escobar PF, Scharm C, et al. Three-dimensional power Doppler ultrasound improves the diagnostic accuracy for ovarian cancer prediction. Gynecol Oncol 2001;82(1):40-8.

34. DePriest PD, Gallion HH, Pavlik EJ, et al. Transvaginal sonography as a screening method for the detection of early ovarian cancer. Gynecol Oncol 1997;65(3):408-14.

35. DePriest PD, Shenson D, Fried A, et al. A morphology index based on sonographic findings in ovarian cancer. Gynecol Oncol 1993;51(1):7-11.

36. DePriest PD, van Nagell JR Jr, Gallion HH, et al. Ovarian cancer screening in asymptomatic postmenopausal women. Gynecol Oncol 1993;51(2):205-9.

37. Dottino PR, Levine DA, Ripley DL, et al. Laparoscopic management of adnexal masses in premenopausal and postmenopausal women. Obstet Gynecol 1999;93(2):223-8.

38. Fleischer AC, Cullinan JA, Jones HW 3rd, et al. Correlation of histomorphology of ovarian masses with color Doppler sonography. Ultrasound Med Biol 1996;22(5):555-9.

39. Lin JY, Angel C, DuBeshter B, et al. Diagnoses after laparotomy for a mass in the pelvic area in women. Surg Gynecol Obstet 1993;176(4):333-8.

40. Modesitt SC, Pavlik EJ, Ueland FR, et al. Risk of malignancy in unilocular ovarian cystic tumors less than 10 centimeters in diameter. Obstet Gynecol 2003;102(3):594-9.

41. Parker WH, Levine RL, Howard FM, et al. A multicenter study of laparoscopic management of selected cystic adnexal masses in postmenopausal women. J Am Coll Surg 1994;179(6):733-7.

42. Roman LD, Muderspach LI, Stein SM, et al. Pelvic examination, tumor marker level, and gray-scale and Doppler sonography in the prediction of pelvic cancer. Obstet Gynecol 1997;89(4):493-500.

43. Schneider VL, Schneider A, Reed KL, et al. Comparison of Doppler with two-dimensional sonography and CA 125 for prediction of malignancy of pelvic masses. Obstet Gynecol 1993;81(6):983-8.

44. Scoutt LM, McCarthy SM, Lange R, et al. MR evaluation of clinically suspected adnexal masses. J Comput Assist Tomogr 1994;18(4):609-18.

45. Shen-Gunther J, Mannel RS. Ascites as a predictor of ovarian malignancy. Gynecol Oncol 2002;87(1):77-83.

46. Smikle CB, Lunt CC, Hankins GD. Clinical predictors in the evaluation of a pelvic mass. Mil Med 1995;160(5):233-5.

47. Troiano RN, Quedens-Case C, Taylor KJ. Correlation of findings on transvaginal sonography with serum CA 125 levels. AJR Am J Roentgenol 1997;168(6):1587-90.

48. Twickler DM, Forte TB, Santos-Ramos R, et al. The Ovarian Tumor Index predicts risk for malignancy. Cancer 1999;86(11):2280-90.

49. van Nagell JR Jr, DePriest PD, Reedy MB, et al. The efficacy of transvaginal sonographic screening in asymptomatic women at risk for ovarian cancer. Gynecol Oncol 2000;77(3):350-6.

50. Vasilev SA, Schlaerth JB, Campeau J, et al. Serum CA 125 levels in preoperative evaluation of pelvic masses. Obstet Gynecol 1988;71(5):751-6.

51. Adonakis GL, Paraskevaidis E, Tsiga S, et al. A combined approach for the early detection of ovarian cancer in asymptomatic women. Eur J Obstet Gynecol Reprod Biol 1996;65(2):221-5.

52. Andolf E, Jorgensen C, Astedt B. Ultrasound examination for detection of ovarian carcinoma in risk groups. Obstet Gynecol 1990;75(1):106-9.

53. Balbi GC, Musone R, Menditto A, et al. Women with a pelvic mass: indicators of malignancy. Eur J Gynaecol Oncol 2001;22(6):459-62.

54. Buckshee K, Temsu I, Bhatla N, et al. Pelvic examination, transvaginal ultrasound and transvaginal color Doppler sonography as predictors of ovarian cancer. Int J Gynaecol Obstet 1998;61(1):51-7.

55. Dowd JR, Quinn MA, Rome R, et al. Women with a pelvic mass--when to perform an ultrasound. Aust N Z J Obstet Gynaecol 1993;33(4):404-7.

56. Finkler NJ, Benacerraf B, Lavin PT, et al. Comparison of serum CA 125, clinical impression, and ultrasound in the preoperative evaluation of ovarian masses. Obstet Gynecol 1988;72(4):659-64.

57. Grover SR, Quinn MA. Is there any value in bimanual pelvic examination as a screening test. Med J Aust 1995;162(8):408-10.

58. Jacobs I, Stabile I, Bridges J, et al. Multimodal approach to screening for ovarian cancer. Lancet 1988;1(8580):268-71.

59. Ong S, Duffy T, Murphy J. Transabdominal ultrasound and its correlation with clinical findings in gynaecology. Ir J Med Sci 1996;165(4):268-70.

60. Padilla LA, Radosevich DM, Milad MP. Accuracy of the pelvic examination in detecting adnexal masses. Obstet Gynecol 2000;96(4):593-8.

61. Padilla LA, Radosevich DM, Milad MP. Limitations of the pelvic examination for evaluation of the female pelvic organs. Int J Gynaecol Obstet 2005;88(1):84-8.

62. Schutter EM, Kenemans P, Sohn C, et al. Diagnostic value of pelvic examination, ultrasound, and serum CA 125 in postmenopausal women with a pelvic mass. An international multicenter study. Cancer 1994;74(4):1398-406.

63. Schutter EM, Sohn C, Kristen P, et al. Estimation of probability of malignancy using a logistic model combining physical examination, ultrasound, serum CA 125, and serum CA 72-4 in postmenopausal women with a pelvic mass: an international multicenter study. Gynecol Oncol 1998;69(1):56-63.

64. Ueland F, DePriest P, DeSimone C, et al. The accuracy and examination under anesthesia and transvaginal sonography in evaluating ovarian size. Gynecol Oncol 2005 (in Press) .

65. Alcazar JL, Castillo G. Comparison of 2-dimensional and 3-dimensional power-Doppler imaging in complex adnexal masses for the prediction of ovarian cancer. Am J Obstet Gynecol 2005;192(3):807-12.

66. Alcazar JL, Errasti T, Zornoza A, et al. Transvaginal color Doppler ultrasonography and CA-125 in suspicious adnexal masses. Int J Gynaecol Obstet 1999;66(3):255-61.

67. Alcazar JL, Galan MJ, Garcia-Manero M, et al. Three-dimensional sonographic morphologic assessment in complex adnexal masses: preliminary experience. J Ultrasound Med 2003;22(3):249-54.

68. Alcazar JL, Lopez-Garcia G. Transvaginal color Doppler assessment of venous flow in adnexal masses. Ultrasound Obstet Gynecol 2001;17(5):434-8.

69. Alcazar JL, Merce LT, Laparte C, et al. A new scoring system to differentiate benign from malignant adnexal masses. Am J Obstet Gynecol 2003;188(3):685-92.

70. Anandakumar C, Chew S , Wong YC , et al. Role of transvaginal ultrasound color flow imaging and Doppler waveform analysis in differentiating between benign and malignant ovarian tumors. Ultrasound Obstet Gynecol 1996;7:280-4.

71. Antonic J, Rakar S. Colour and pulsed Doppler US and tumour marker CA 125 in differentiation between benign and malignant ovarian masses. Anticancer Res 1995;15:1527-32.

72. Asif N, Sattar A, Dawood MM, et al. Pre-operative evaluation of ovarian mass: risk of malignancy index. J Coll Physicians Surg Pak 2004;14(3):128-31.

73. Benjapibal M, Sunsaneevitayakul P, Boriboonhirunsarn D, et al. Color Doppler ultrasonography for prediction of malignant ovarian tumors. J Med Assoc Thai 2002;85(6):709-15.

74. Benjapibal M, Sunsaneevitayakul P, Phatihattakorn C, et al. Sonographic morphological pattern in the pre-operative prediction of ovarian masses. J Med Assoc Thai 2003;86(4):332-7.

75. Berlanda N, Ferrari MM, Mezzopane R, et al. Impact of a multiparameter, ultrasound-based triage on surgical management of adnexal masses. Ultrasound Obstet Gynecol 2002;20(2):181-5.

76. Bromley B, Goodman H, Benacerraf BR. Comparison between sonographic morphology and Doppler waveform for the diagnosis of ovarian malignancy. Obstet Gynecol 1994;83(3):434-7.

77. Brown DL, Doubilet PM, Miller FH, et al. Benign and malignant ovarian masses: selection of the most discriminating gray-scale and Doppler sonographic features. Radiology 1998;208(1):103-10.

78. Buist MR, Golding RP, Burger CW, et al. Comparative evaluation of diagnostic methods in ovarian carcinoma with emphasis on CT and MRI. Gynecol Oncol 1994;52(2):191-8.

79. Buy JN, Ghossain MA, Hugol D, et al. Characterization of adnexal masses: combination of color Doppler and conventional sonography compared with spectral Doppler analysis alone and conventional sonography alone. AJR Am J Roentgenol 1996;166(2):385-93.

80. Canis M, Pouly JL, Wattiez A, et al. Laparoscopic management of adnexal masses suspicious at ultrasound. Obstet Gynecol 1997;89(5 Pt 1):679-83.

81. Carter JR, Lau M, Fowler JM, et al. Blood flow characteristics of ovarian tumors: implications for ovarian cancer screening. Am J Obstet Gynecol 1995;172(3):901-7.

82. Carter PG, Iles RK, Neven P, et al. The measurement of urinary beta core fragment in conjunction with serum CA125 does not aid the differentiation of malignant from benign pelvic masses. Gynecol Oncol 1993;51(3):368-71.

83. Caruso A, Caforio L, Testa AC, et al. Transvaginal color Doppler ultrasonography in the presurgical characterization of adnexal masses. Gynecol Oncol 1996;63(2):184-91.

84. Chen DX, Schwartz PE, Li FQ. Saliva and serum CA 125 assays for detecting malignant ovarian tumors. Obstet Gynecol 1990;75(4):701-4.

85. Chen DX, Schwartz PE, Li XG, et al. Evaluation of CA 125 levels in differentiating malignant from benign tumors in patients with pelvic masses. Obstet Gynecol 1988;72(1):23-7.

86. Chou CY, Chang CH, Yao BL , et al. Color Doppler ultrasonography and serum CA 125 in the differentiation of benign and malignant ovarian tumors. J Clin Ultrasound 1994;22:491-6.

87. Davies AP, Jacobs I, Woolas R, et al. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. Br J Obstet Gynaecol 1993;100(10):927-31.

88. DePriest PD, Varner E, Powell J. The efficacy of a sonographic morphology index in identifying ovarian cancer: a multi-institutional investigation. Gynecol Oncol 1994;55:174-8.

89. Einhorn N, Bast RC Jr, Knapp RC, et al. Preoperative evaluation of serum CA 125 levels in patients with primary epithelial ovarian cancer. Obstet Gynecol 1986;67(3):414-6.

90. Ekerhovd E, Wienerroith H, Staudach A, et al. Preoperative assessment of unilocular adnexal cysts by transvaginal ultrasonography: a comparison between ultrasonographic morphologic imaging and histopathologic diagnosis. Am J Obstet Gynecol. 2001;184(2):48-54.

91. Fenchel S, Grab D, Nuessle K, et al. Asymptomatic adnexal masses: correlation of FDG PET and histopathologic findings. Radiology 2002;223(3):780-8.

92. Ferdeghini M, Gadducci A, Prontera C, et al. Serum soluble interleukin-2 receptor assay in epithelial ovarian cancer. Tumour Biol 1993;14(5):303-9.

93. Ferrazzi E, Zanetta G, Dordoni D, et al. Transvaginal ultrasonographic characterization of ovarian masses: comparison of five scoring systems in a multicenter study.[erratum appears in Ultrasound Obstet Gynecol 1998 Jan;11(1):v Note: Lissoni G[corrected to Lissoni AA]]. Ultrasound Obstet Gynecol 1997;10(3):192-7.

94. Fleischer AC, Rodgers WH, Kepple DM, et al. Color Doppler sonography of benign and malignant ovarian masses. Radiographics 1992;12(5):879-85.

95. Franchi M, Beretta P, Ghezzi F, et al. Diagnosis of pelvic masses with transabdominal color Doppler, CA 125 and ultrasonography. Acta Obstet Gynecol Scand 1995;74(9):734-9.

96. Gadducci A, Baicchi U, Marrai R, et al. Preoperative evaluation of D-dimer and CA 125 levels in differentiating benign from malignant ovarian masses. Gynecol Oncol 1996;60(2):197-202.

97. Gadducci A, Capriello P, Bartolini T, et al. The association of ultrasonography and CA-125 test in the preoperative evaluation of ovarian carcinoma. Eur J Gynaecol Oncol 1988;9(5):373-6.

98. Gadducci A, Ferdeghini M, Prontera C, et al. The concomitant determination of different tumor markers in patients with epithelial ovarian cancer and benign ovarian masses: relevance for differential diagnosis. Gynecol Oncol 1992;44(2):147-54.

99. Gadducci A, Ferdeghini M, Rispoli G, et al. Comparison of tumor-associated trypsin inhibitor (TATI) with CA125 as a marker for diagnosis and monitoring of epithelial ovarian cancer. Scand J Clin Lab Invest Suppl 1991;207(19-24).

100. Grab D, Flock F, Stohr I, et al. Classification of asymptomatic adnexal masses by ultrasound, magnetic resonance imaging, and positron emission tomography. Gynecol Oncol 2000;77(3):454-9.

101. Granberg S, Norstrom A, Wikland M. Tumors in the lower pelvis as imaged by vaginal sonography. Gynecol Oncol 1990;37(2):224-9.

102. Granberg S, Norstrom A, Wikland M. Comparison ofendovaginal ultrasound and cytological evaluation of cystic ovarian tumors. J Ultrasound Med 1991;10:9-14.

103. Guerriero S, Ajossa S, Risalvato A, et al. Diagnosis of adnexal malignancies by using color Doppler energy imaging as a secondary test in persistent masses. Ultrasound Obstet Gynecol 1998;11(4):277-82.

104. Guerriero S, Ajossa S, Garau N, et al. Ultrasonography and color Doppler-based triage for adnexal masses to provide the most appropriate surgical approach. Am J Obstet Gynecol 2005;192(2):401-6.

105. Guerriero S, Alcazar JL, Coccia ME, et al. Complex pelvic mass as a target of evaluation of vessel distribution by color Doppler sonography for the diagnosis of adnexal malignancies: results of a multicenter European study. J Ultrasound Med 2002;21(10):1105-11.

106. Hata K, Hata T, Manabe A, et al. A critical evaluation of transvaginal Doppler studies, transvaginal sonography, magnetic resonance imaging, and CA 125 in detecting ovarian cancer. Obstet Gynecol 1992;80(6):922-6.

107. Hata K, Hata T, Kitao M. Intratumoral peak systolic velocity as a new possible predictor for detection of adnexal malignancy. Am J Obstet Gynecol 1995;172:1496-500.

108. Herrmann UJ Jr, Locher GW, Goldhirsch A. Sonographic patterns of ovarian tumors: prediction of malignancy. Obstet Gynecol 1987;69(5):777-81.

109. Hillaby K, Aslam N, Salim R, et al. The value of detection of normal ovarian tissue (the 'ovarian crescent sign') in the differential diagnosis of adnexal masses. Ultrasound Obstet Gynecol 2004;23(1):63-7.

110. Hogdall EV, Hogdall CK, Tingulstad S, et al. Predictive values of serum tumour markers tetranectin, OVX1, CASA and CA125 in patients with a pelvic mass. Int J Cancer 2000;89(6):519-23.

111. Hricak H, Chen M, Coakley FV, et al. Complex adnexal masses: detection and characterization with MR imaging--multivariate analysis. Radiology 2000;214(1):39-46.

112. Huber S, Medl M, Baumann L, et al. Value of ultrasound and magnetic resonance imaging in the preoperative evaluation of suspected ovarian masses. Anticancer Res 2002;22(4):2501-7.

113. Hurteau JA, Woolas RP, Jacobs IJ, et al. Soluble interleukin-2 receptor alpha is elevated in sera of patients with benign ovarian neoplasms and epithelial ovarian cancer. Cancer 1995;76(9):1615-20.

114. Inoue M, Fujita M, Nakazawa A, et al. Sialyl-Tn, sialyl-Lewis Xi, CA 19-9, CA 125, carcinoembryonic antigen, and tissue polypeptide antigen in differentiating ovarian cancer from benign tumors. Obstet Gynecol 1992;79(3):434-40.

115. Itakura T, Kikkawa F, Kajiyama H, et al. Doppler flow and arterial location in ovarian tumors. Int J Gynaecol Obstet 2003;83(3):277-83.

116. Jacobs I, Oram D, Fairbanks J, et al. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. Br J Obstet Gynaecol 1990;97(10):922-9.

117. Jain KA. Prospective evaluation of adnexal masses with endovaginal gray-scale and duplex and color Doppler US: correlation with pathologic findings. Radiology 1994;191(1):63-7.

118. Jain KA, Friedman DL, Pettinger TW, et al. Adnexal masses: comparison of specificity of endovaginal US and pelvic MR imaging. Radiology 1993;186(3):697-704.

119. Juhasz B, Kurjak A, Lampe L, et al. Tissue characterization by transvaginal colour doppler for the evaluation of gynaecological tumours. Acta Med Hung 1990;47:149-56.

120. Kawai M, Kikkawa F, Ishikawa H. Differential diagnosis of ovarian tumors by transvaginal color - pulse Doppler sonography. Gynecol Oncol 1994;54:209-14.

121. Kawahara K, Yoshida Y, Kurokawa T, et al. Evaluation of positron emission tomography with tracer 18-fluorodeoxyglucose in addition to magnetic resonance imaging in the diagnosis of ovarian cancer in selected women after ultrasonography. J Comput Assist Tomogr 2004;28(4):505-16.

122. Komatsu T, Konishi I, Mandai M, et al. Adnexal masses: transvaginal US and gadolinium-enhanced MR imaging assessment of intratumoral structure. Radiology 1996;198(1):109-15.

123. Kurjak A, Kupesic S. Three dimensional ultrasound and power doppler in assessment of uterine and ovarian angiogenesis: a prospective study. Croat Med J 1999;40(3):413-20.

124. Kurjak A, Kupesic S, Sparac V, et al. Three-dimensional ultrasonographic and power Doppler characterization of ovarian lesions. Ultrasound Obstet Gynecol 2000;16(4):365-71.

125. Kurjak A, Predanic M. New scoring system for prediction of ovarian malignancy based on transvaginal color Doppler sonography. J Ultrasound Med 1992;11(12):631-8.

126. Kurjak A, Schulman H, Sosic A, et al. Transvaginal ultrasound, color flow, and Doppler waveform of the postmenopausal adnexal mass. Obstet Gynecol 1992;80(6):917-21.

127. Kurjak A, Shalan H, Kupesic S, et al. An attempt to screen asymptomatic women for ovarian and endometrial cancer with transvaginal color and pulsed Doppler sonography. J Ultrasound Med 1994;13(4):295-301.

128. Kurjak A, Zalud I, Alfirevic Z. Evaluation of adnexal masses with transvaginal color ultrasound. J Ultrasound Med 1991;10(6):295-7.

129. Kurtz AB, Tsimikas JV, Tempany CM, et al. Diagnosis and staging of ovarian cancer: comparative values of Doppler and conventional US, CT, and MR imaging correlated with surgery and histopathologic analysis--report of the Radiology Diagnostic Oncology Group. Radiology 1999;212(1):19-27.

130. Leeners B, Schild RL, Funk A, et al. Colour Doppler sonography improves the pre-operative diagnosis of ovarian tumours made using conventional transvaginal sonography. Eur J Obstet Gynecol Reprod Biol 1996;64(1):79-85.

131. Lerner JP, Timor-Tritsch IE, Federman A, et al. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. Am J Obstet Gynecol 1994;170(1 Pt 1):81-5.

132. Lin P, Wu C, Lee C, et al. Color Doppler Ultrasound in the Assessment of Ovarian Neoplasms. J Med Ultrasound 1993;1(4):172-6.

133. Luxman D, Bergman A, Sagi J, et al. The postmenopausal adnexal mass: correlation between ultrasonic and pathologic findings. Obstet Gynecol 1991;77(5):726-8.

134. Ma S, Shen K, Lang J. A risk of malignancy index in preoperative diagnosis of ovarian cancer. Chin Med J 2003;116(3):396-9.

135. Maggino T, Gadducci A, D'Addario V, et al. Prospective multicenter study on CA 125 in postmenopausal pelvic masses. Gynecol Oncol 1994;54(2):117-23.

136. Malkasian GD Jr, Knapp RC, Lavin PT, et al. Preoperative evaluation of serum CA 125 levels in premenopausal and postmenopausal patients with pelvic masses: discrimination of benign from malignant disease. Am J Obstet Gynecol 1988;159(2):341-6.

137. Maly Z, Riss P, Deutinger J. Localization of blood vessels and qualitative assessment of blood flow in ovarian tumors. Obstet Gynecol 1995;85(1):33-6.

138. Mancuso A, De Vivo A, Triolo O, et al. The role of transvaginal ultrasonography and serum CA 125 assay combined with age and hormonal state in the differential diagnosis of pelvic masses. Eur J Gynaecol Oncol 2004;25(2):207-10.

139. Manjunath AP, Pratapkumar, Sujatha K, et al. Comparison of three risk of malignancy indices in evaluation of pelvic masses. Gynecol Oncol 2001;81(2):225-9.

140. Marchetti M, Zambon A, Lamaina V, et al. Ultrasound as a possible screening method in ovarian cancer.[erratum appears in Eur J Gynaecol Oncol. 2002;23(3):172 Note: Marchioro S [corrected to Marchiori S]]. Eur J Gynaecol Oncol 2002;23(2):123-6.

141. Marret H, Sauget S, Giraudeau B, et al. Contrast-enhanced sonography helps in discrimination of benign from malignant adnexal masses. J Ultrasound Med. 2004;23(12):1629-39; quiz 1641-42.

142. Matthes AC, Moreira de Andrade JM , Bighetti S. Selection of criteria for the treatment of ovarian cysts on the bases of ultrasound and cytology. Gynecol Obstet Invest 1996;42:244-8.

143. McIntosh MW, Drescher C, Karlan B, et al. Combining CA125 and SMR serum markers for diagnosis and early detection of ovarian cancer. Gynecol Oncol 2004;95:9-15.

144. Medl M, Kulenkampff KJ, Stiskal M, et al. Magnetic resonance imaging in the preoperative evaluation of suspected ovarian masses. Anticancer Res 1995;15(3):1123-5.

145. Menon U, Talaat A, Rosenthal AN, et al. Performance of ultrasound as a second line test to serum CA125 in ovarian cancer screening. BJOG 2000;107(2):165-9.

146. Merce LT, Caballero RA, Barco MJ, et al. B-mode, utero-ovarian and intratumoural transvaginal colour Doppler ultrasonography for differential diagnosis of ovarian tumours. Eur J Obstet Gynecol Reprod Biol 1998;76(1):97-107.

147. Morgante G, la Marca A, Ditto A, et al. Comparison of two malignancy risk indices based on serum CA125, ultrasound score and menopausal status in the diagnosis of ovarian masses. Br J Obstet Gynaecol 1999;106(6):524-7.

148. O'Connell GJ, Ryan E, Murphy KJ, et al. Predictive value of CA 125 for ovarian carcinoma in patients presenting with pelvic masses. Obstet Gynecol 1987;70(6):930-2.

149. Onsrud M, Shabana A, Austgulen R. Soluble tumor necrosis factor receptors and CA 125 in serum as markers for epithelial ovarian cancer. Tumour Biol 1996;17(2):90-6.

150. Padungsutt P, Thirapagawong C, Senapad S, et al. Accuracy of tissue polypeptide specific antigen (TPS) in the diagnosis of ovarian malignancy. Anticancer Res 2000;20(2B):1291-5.

151. Patsner B, Mann WJ. The value of preoperative serum CA 125 levels in patients with a pelvic mass. Am J Obstet Gynecol 1988;159(4):873-6.

152. Prompeler HJ, Madjar H, Sauerbrei W. Classification of adnexal tumors by transvaginal color Doppler. Gynecol Oncol 1996;61(3):354-63.

153. Pyrgiotis E, Salamalekis E, Loghis C, et al. CA 125 in preoperative evaluation of pelvic masses. Eur J Gynaecol Oncol 1993;14(4):279-82.

154. Rehn M, Lohmann K, Rempen A. Transvaginal ultrasonography of pelvic masses: evaluation of B-mode technique and Doppler ultrasonography. Am J Obstet Gynecol 1996;175(1):97-104.

155. Reles A, Wein U, Lichtenegger W. Transvaginal color Doppler sonography and conventional sonography in the preoperative assessment of adnexal masses. J Clin Ultrasound 1997;25(5):217-25.

156. Reuter M, Steffens J. , Schuppler U. Critical evaluation of the specificity of MRI and TVUs for differentiation of malignant from benign adnexal lesions. Eur Radiol 1998;8(39-44).

157. Roman LD, Muderspach LI, Burnett AF, et al. Carcinoembryonic antigen in women with isolated pelvic masses. Clinical utility? J Reprod Med 1998;43(5):403-7.

158. Salem S, White LM, Lai J. Doppler sonography of adnexal masses: the predictive value of the pulsatility index in benign and malignant disease. AJR Am J Roentgenol 1994;163(5):1147-50.

159. Sassone AM, Timor-Tritsch IE, Artner A, et al. Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy. Obstet Gynecol 1991;78(1):70-6.

160. Sawicki W, Spiewankiewicz B, Cendrowski K, et al. Preoperative discrimination between malignant and benign adnexal masses with transvaginal ultrasonography and colour blood flow imaging. Eur J Gynaecol Oncol 2001;22(2):137-42.

161. Schelling M, Braun M, Kuhn W, et al. Combined transvaginal B-mode and color Doppler sonography for differential diagnosis of ovarian tumors: results of a multivariate logistic regression analysis. Gynecol Oncol 2000;77(1):78-86.

162. Schutter EM, Davelaar EM, van Kamp GJ, et al. The differential diagnostic potential of a panel of tumor markers (CA 125, CA 15-3, and CA 72-4 antigens) in patients with a pelvic mass. Am J Obstet Gynecol 2002;187(2):385-92.

163. Sengoku K, Satoh T, Saitoh S, et al. Evaluation of transvaginal color Doppler sonography, transvaginal sonography and CA 125 for prediction of ovarian malignancy. Int J Gynaecol Obstet 1994;46(1):39-43.

164. Shabana A, Onsrud M. Tissue polypeptide-specific antigen and CA 125 as serum tumor markers in ovarian carcinoma. Tumour Biol 1994;15(6):361-7.

165. Siegel BA, Dehdashti F, Mutch DG, et al. Evaluation of 111In-DTPA-folate as a receptor-targeted diagnostic agent for ovarian cancer: initial clinical results. J Nucl Med 2003;44(5):700-7.

166. Sohaib SA, Mills TD, Sahdev A, et al. The role of magnetic resonance imaging and ultrasound in patients with adnexal masses. Clin Radiol 2005;60(3):340-8.

167. Soper JT, Hunter VJ, Daly L, et al. Preoperative serum tumor-associated antigen levels in women with pelvic masses. Obstet Gynecol 1990;75(2):249-54.

168. Stein SM, Laifer-Narin S, Johnson MB, et al. Differentiation of benign and malignant adnexal masses: relative value of gray-scale, color Doppler, and spectral Doppler sonography. AJR Am J Roentgenol 1995;164(2):381-6.

169. Strigini FA, Gadducci A, Del Bravo B, et al. Differential diagnosis of adnexal masses with transvaginal sonography, color flow imaging, and serum CA 125 assay in pre- and postmenopausal women. Gynecol Oncol 1996;61(1):68-72.

170. Szpurek D, Moszyniki R, Sajdak S. Clinical value of the ultrasound Doppler index in determination of ovarian tumor malignancy. Eur J Gynaecol Oncol 2004;25(4):442-4.

171. Tailor A, Bourne TH, Campbell S, et al. Results from an ultrasound-based familial ovarian cancer screening clinic: a 10-year observational study. Ultrasound Obstet Gynecol 2003;21(4):378-85.

172. Takac I. Analysis of blood flow in adnexal tumors by using color Doppler imaging and pulsed spectral analysis. Ultrasound Med Biol 1998;24(8):1137-42.

173. Tanir HM, Ozalp S, Yalcin OT, et al. Preoperative serum vascular endothelial growth factor (VEGF) in ovarian masses. Eur J Gynaecol Oncol 2003;24(3-4):271-4.

174. Tay SK, Chua EK. Correlation of serum, urinary and salivary CA 125 levels in patients with adnexal masses. Ann Acad Med Singapore 1994;23(3):311-4.

175. Tekay A, Jouppila P. Validity of pulsatility and resistance indices in classification of adnexal tumors with transvaginal color doppler ultrasound. Ultrasound Obstet Gynecol 1992;2:338-44.

176. Tepper R, Lerner-Geva L, Altaras MM, et al. Transvaginal color flow imaging in the diagnosis of ovarian tumors. J Ultrasound Med 1995;14(10):731-4.

177. Tian J, Zhang J, Jiao L, et al. A prospective study of Tc-99m MIBI in the differential diagnosis of pelvic masses in female patients. Clin Nucl Med 2000;25(8):614-8.

178. Timmerman D, Verrelst H, Bourne TH, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound Obstet Gynecol 1999;13(1):17-25.

179. Timor-Tritsch LE, Lerner J, Monteagudo A, et al. Transvaginal ultrasonographic characterization of ovarian masses by means of color flow-directed Doppler measurements and a morphologic scoring system. Am J Obstet Gynecol 1993;168:909-13.

180. Tingulstad S, Hagen B, Skjeldestad FE, et al. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. Br J Obstet Gynaecol 1996;103(8):826-31.

181. Torres JC, Derchain SF, Faundes A, et al. Risk-of-malignancy index in preoperative evaluation of clinically restricted ovarian cancer. Rev Paul Med 2002;120(3):72-6.

182. Valentin L. Gray scale sonography, subjective evaluation of the color Doppler image and measurement of blood flow velocity for distinguishing benign and malignant tumors of suspected adnexal origin. Eur J Obstet Gynecol Reprod Biol 1997;72(1):63-72.

183. Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. Ultrasound Obstet Gynecol 1999;14(5):338-47.

184. Valentin L. Comparison of Lerner score, Doppler ultrasound examination, and their combination for discrimination between benign and malignant adnexal masses. Ultrasound Obstet Gynecol 2000;15(2):143-7.

185. Valentin L, Hagen B, Tingulstad S, et al. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. Ultrasound Obstet Gynecol 2001;18(4):357-65.

186. Vuento MH, Pirhonen JP, Makinen JI, et al. Evaluation of ovarian findings in asymptomatic postmenopausal women with color Doppler ultrasound. Cancer 1995;76(7):1214-8.

187. Wakahara F, Kikkawa F, Nawa A, et al. Diagnostic efficacy of tumor markers, sonography, and intraoperative frozen section for ovarian tumors. Gynecol Obstet Invest 2001;52(3):147-52.

188. Weiner Z, Thaler I, Beck D, et al. Differentiating malignant from benign ovarian tumors with transvaginal color flow imaging. Obstet Gynecol 1992;79(2):159-62.

189. Woolas RP, Conaway MR, Xu F, et al. Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. Gynecol Oncol 1995;59(1):111-6.

190. Wu CC, Lee CN, Chen T, et al. Factors contributing to the accuracy in diagnosing ovarian malignancy by color Doppler ultrasound. Obstet Gynecol 1994;84:605-8.

191. Yamashita Y, Hatanaka Y, Torashima M, et al. Characterization of sonographically indeterminate ovarian tumors with MR imaging. A logistic regression analysis. Acta Radiol 1997;38(4 Pt 1):572-7.

192. Yamashita Y, Torashima M, Hatanaka Y, et al. Adnexal masses: accuracy of characterization with transvaginal US and precontrast and postcontrast MR imaging. Radiology 1995;194(2):557-65.

193. Zanetta G, Vergani P, Lissoni A. Color Doppler ultrasound in the preoperative assessment of adnexal masses. Acta Obstet Gynecol Scand 1994;73:637-41.

194. Zhang Z, Barnhill SD, Zhang H, et al. Combination of multiple serum markers using an artificial neural network to improve specificity in discriminating malignant from benign pelvic masses. Gynecol Oncol 1999;73(1):56-61.

195. Zimmer Y, Tepper R, Akselrod S. An automatic approach for morphological analysis and malignancy evaluation of ovarian masses using B-scans. Ultrasound Med Biol 2003;29(11):1561-70.

196. Timmerman D, Schwarzler P, Collins WP, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol 1999;13(1):11-6.

197. Kinkel K, Hricak H, Lu Y, et al. US characterization of ovarian masses: a meta-analysis. Radiology 2000;217(3):803-11.

198. Larsen T, Torp-Pedersen ST, Ottesen M, et al. Abdominal ultrasound combined with histological and cytological fine needle biopsy of suspected ovarian tumors. Eur J Obstet Gynecol Reprod Biol 1993;50(3):203-9.

199. Alcazar JL, Ruiz-Perez ML, Errasti T. Transvaginal color Doppler sonography in adnexal masses: which parameter performs best? Ultrasound Obstet Gynecol 1996;8(2):114-9.

200. Kurjak A, Shalan H, Kupesic S, et al. Transvaginal color Doppler sonography in the assessment of pelvic tumor vascularity. Ultrasound Obstet Gynecol 1993;3(2):137-54.

201. Valentin L. Prospective cross-validation of Doppler ultrasound examination and gray-scale ultrasound imaging for discrimination of benign and malignant pelvic masses. Ultrasound Obstet Gynecol 1999;14(4):273-83.

202. Andersen ES, Knudsen A, Rix P, et al. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. Gynecol Oncol 2003;90(1):109-12.

203. Aslam N, Banerjee S, Carr JV, et al. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. Obstet Gynecol 2000;96(1):75-80.

204. Aslam N, Tailor A, Lawton F, et al. Prospective evaluation of three different models for the pre-operative diagnosis of ovarian cancer. BJOG 2000;107(11):1347-53.

205. Biagiotti R, Desii C, Vanzi E, et al. Predicting ovarian malignancy: application of artificial neural networks to transvaginal and color Doppler flow US. Radiology 1999;210(2):399-403.

206. Lu C, Van Gestel T, Suykens JA, et al. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. Artif Intell Med 2003;28(3):281-306.

207. Mol BW, Boll D, De Kanter M, et al. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. Gynecol Oncol 2001;80(2):162-7.

208. Obeidat BR, Amarin ZO, Latimer JA, et al. Risk of malignancy index in the preoperative evaluation of pelvic masses. Int J Gynaecol Obstet 2004;85(3):255-8.

209. Tailor A, Jurkovic D, Bourne TH, et al. Sonographic prediction of malignancy in adnexal masses using an artificial neural network. Br J Obstet Gynaecol 1999;106(1):21-30.

210. Timmerman D, Bourne TH, Tailor A, et al. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model. Am J Obstet Gynecol 1999;181(1):57-65.

211. Tingulstad S, Hagen B, Skjeldestad FE, et al. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. Obstet Gynecol 1999;93(3):448-52.

212. Jacobs IJ, Rivera H, Oram DH, et al. Differential diagnosis of ovarian cancer with tumour markers CA 125, CA 15-3 and TAG 72.3. Br J Obstet Gynaecol 1993;100(12):1120-4.

213. Alcazar JL, Jurado M. Using a logistic model to predict malignancy of adnexal masses based on menopausal status, ultrasound morphology, and color Doppler findings. Gynecol Oncol 1998;69(2):146-50.

214. Castillo G, Alcazar JL, Jurado M. Natural history of sonographically detected simple unilocular adnexal cysts in asymptomatic postmenopausal women. Gynecol Oncol 2004;92(3):965-9.

215. Goldstein SR, Subramanyam B, Snyder JR, et al. The postmenopausal cystic adnexal mass: the potential role of ultrasound in conservative management. Obstet Gynecol 1989;73(1):8-10.

216. Levine D, Gosink BB, Wolf SI, et al. Simple adnexal cysts: the natural history in postmenopausal women.[see comment]. Radiology 1992;184(3):653-9.

217. Schincaglia P, Brondelli L, Cicognani A, et al. A feasibility study of ovarian cancer screening: does fine-needle aspiration improve ultrasound specificity? Tumori 1994;80(3):181-7.

218. Valentin L, Akrawi D. The natural history of adnexal cysts incidentally detected at transvaginal ultrasound examination in postmenopausal women. Ultrasound Obstet Gynecol 2002;20(2):174-80.

219. Canis M, Mage G, Pouly JL, et al. Laparoscopic diagnosis of adnexal cystic masses: a 12-year experience with long-term follow-up. Obstet Gynecol 1994;83(5 Pt 1):707-12.

220. Canis M, Mashiach R, Wattiez A, et al. Frozen section in laparoscopic management of macroscopically suspicious ovarian masses. J Am Assoc Gynecol Laparosc 2004;11(3):365-9.

221. Carley ME, Klingele CJ, Gebhart JB, et al. Laparoscopy versus laparotomy in the management of benign unilateral adnexal masses. J Am Assoc Gynecol Laparosc 2002;9(3):321-6.

222. Chapron C, Dubuisson JB, Capella-Allouc S. Salpingo-oophorectomy for adnexal masses. Place and results for operative laparoscopy. Eur J Obstet Gynecol Reprod Biol 1997;73(1):43-8.

223. Chi DS, Abu-Rustum NR, Sonoda Y, et al. Ten-year experience with laparoscopy on a gynecologic oncology service: analysis of risk factors for complications and conversion to laparotomy. Am J Obstet Gynecol 2004;191(4):1138-45.

224. Deckardt R, Saks M, Graeff H. Comparison of minimally invasive surgery and laparotomy in the treatment of adnexal masses. J Am Assoc Gynecol Laparosc 1994;1(4 Pt 1):333-8.

225. Fanfani F, Fagotti A, Ercoli A, et al. A prospective randomized study of laparoscopy and minilaparotomy in the management of benign adnexal masses. Hum Reprod 2004;19(10):2367-71.

226. Havrilesky LJ, Peterson BL, Dryden DK, et al. Predictors of clinical outcomes in the laparoscopic management of adnexal masses. Obstet Gynecol 2003;102(2):243-51.

227. Hidlebaugh DA, Vulgaropulos S, Orr RK. Treating adnexal masses. Operative laparoscopy vs. laparotomy. J Reprod Med 1997;42(9):551-8.

228. Lok IH, Sahota DS, Rogers MS, et al. Complications of laparoscopic surgery for benign ovarian cysts. J Am Assoc Gynecol Laparosc 2000;7(4):529-34.

229. Mann WJ, Reich H. Laparoscopic adnexectomy in postmenopausal women. J Reprod Med 1992;37(3):254-6.

230. Marana R, Muzii L, Catalano GF, et al. Laparoscopic excision of adnexal masses. J Am Assoc Gynecol Laparosc 2004;11(2):162-6.

231. Parker J, Proietto AM. Surgical treatment of adnexal cysts by laparotomy. Int J Gynaecol Obstet 1997;59(3):257-8.

232. Sadik S, Onoglu AS, Gokdeniz R, et al. Laparoscopic management of selected adnexal masses. J Am Assoc Gynecol Laparosc 1999;6(3):313-6.

233. Serur E, Emeney PL, Byrne DW. Laparoscopic management of adnexal masses. J Soc Laparoendosc Surg 2001;5(2):143-51.

234. Shalev E, Eliyahu S, Peleg D, et al. Laparoscopic management of adnexal cystic masses in postmenopausal women. Obstet Gynecol 1994;83(4):594-6.

235. Somigliana E, Ragni G, Benedetti F, et al. Does laparoscopic excision of endometriotic ovarian cysts significantly affect ovarian reserve? Insights from IVF cycles. Hum Reprod 2003;18(11):2450-3.

236. Tangjitgamol S, Jesadapatrakul S, Manusirivithaya S, et al. Accuracy of frozen section in diagnosis of ovarian mass. Int J Gynecol Cancer 2004;14(2):212-9.

237. Tarik A, Fehmi C. Complications of gynaecological laparoscopy--a retrospective analysis of 3572 cases from a single institute. J Obstet Gynaecol 2004;24(7):813-6.

238. van Herendael B, Beretta P, Slangen T, et al. Management of adnexal masses by operative laparoscopy. J Am Assoc Gynecol Laparosc 1995;2(3):273-7.

239. Yuen PM, Yu KM, Yip SK, et al. A randomized prospective study of laparoscopy and laparotomy in the management of benign ovarian masses. Am J Obstet Gynecol 1997;177(1):109-14.

240. Myers ER, McCrory DC, Nanda K, et al. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. Am J Epidemiol 2000;151(12):1158-71.

241. Goldie SJ, Kim JJ, Wright TC. Cost-effectiveness of human papillomavirus DNA testing for cervical cancer screening in women aged 30 years or more. Obstet Gynecol 2004;103(4):619-31.

242. Schapira MM, Matchar DB, Young MJ. The effectiveness of ovarian cancer screening. A decision analysis model. Ann Intern Med 1993;118(11):838-43.

243. Urban N, Drescher C, Etzioni R, et al. Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening. Control Clin Trials 1997;18(3):251-70.

244. Skates SJ, Singer DE. Quantifying the potential benefit of CA 125 screening for ovarian cancer. J Clin Epidemiol 1991;44(4-5):365-80.

245. Tengs TO, Winer EP, Paddock S, et al. Testing for the BRCA1 and BRCA2 breast-ovarian cancer susceptibility genes: a decision analysis. Med Decis Making 1998;18(4):365-75.

246. Bast RC Jr, Urban N, Shridhar V, et al. Early detection of ovarian cancer: promise and reality. Cancer Treat Res 2002;107:61-97.

247. Im S, Gordon A, Buttin B, et al. Validation of referral guidelines for women with pelvic masses. Obstet Gyencol 2005;205(1):35-41.

Listing of Included Studies (alphabetical)

Adonakis GL, Paraskevaidis E, Tsiga S, et al. A combined approach for the early detection of ovarian cancer in asymptomatic women. Eur J Obstet Gynecol Reprod Biol 1996;65(2):221-5.

Alcazar JL, Castillo G. Comparison of 2-dimensional and 3-dimensional power-Doppler imaging in complex adnexal masses for the prediction of ovarian cancer. Am J Obstet Gynecol 2005;192(3):807-12.

Alcazar JL, Errasti T, Zornoza A, et al. Transvaginal color Doppler ultrasonography and CA-125 in suspicious adnexal masses. Int J Gynaecol Obstet 1999;66(3):255-61.

Alcazar JL, Galan MJ, Garcia-Manero M, et al. Three-dimensional sonographic morphologic assessment in complex adnexal masses: preliminary experience. J Ultrasound Med 2003;22(3):249-54.

Alcazar JL, Lopez-Garcia G. Transvaginal color Doppler assessment of venous flow in adnexal masses. Ultrasound Obstet Gynecol 2001;17(5):434-8.

Alcazar JL, Merce LT, Laparte C, et al. A new scoring system to differentiate benign from malignant adnexal masses. Am J Obstet Gynecol 2003;188(3):685-92.

Anandakumar C, Chew S , Wong YC , et al. Role of transvaginal ultrasound color flow imaging and Doppler waveform analysis in differentiating between benign and malignant ovarian tumors. Ultrasound Obstet Gynecol 1996;7:280-4.

Andersen ES, Knudsen A, Rix P, et al. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. Gynecol Oncol 2003;90(1):109-12.

Andolf E, Jorgensen C, Astedt B. Ultrasound examination for detection of ovarian carcinoma in risk groups. Obstet Gynecol 1990;75(1):106-9.

Antonic J, Rakar S. Colour and pulsed Doppler US and tumour marker CA 125 in differentiation between benign and malignant ovarian masses. Anticancer Res 1995;15:1527-32.

Asif N, Sattar A, Dawood MM, et al. Pre-operative evaluation of ovarian mass: risk of malignancy index. J Coll Physicians Surg Pak 2004;14(3):128-31.

Aslam N, Banerjee S, Carr JV, et al. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. Obstet Gynecol 2000;96(1):75-80.

Aslam N, Tailor A, Lawton F, et al. Prospective evaluation of three different models for the pre-operative diagnosis of ovarian cancer. BJOG 2000;107(11):1347-53.

Balbi GC, Musone R, Menditto A, et al. Women with a pelvic mass: indicators of malignancy. Eur J Gynaecol Oncol 2001;22(6):459-62.

Benjapibal M, Sunsaneevitayakul P, Boriboonhirunsarn D, et al. Color Doppler ultrasonography for prediction of malignant ovarian tumors. J Med Assoc Thai 2002;85(6):709-15.

Benjapibal M, Sunsaneevitayakul P, Phatihattakorn C, et al. Sonographic morphological pattern in the pre-operative prediction of ovarian masses. J Med Assoc Thai 2003;86(4):332-7.

Berlanda N, Ferrari MM, Mezzopane R, et al. Impact of a multiparameter, ultrasound-based triage on surgical management of adnexal masses. Ultrasound Obstet Gynecol 2002;20(2):181-5.

Biagiotti R, Desii C, Vanzi E, et al. Predicting ovarian malignancy: application of artificial neural networks to transvaginal and color Doppler flow US. Radiology 1999;210(2):399-403.

Bromley B, Goodman H, Benacerraf BR. Comparison between sonographic morphology and Doppler waveform for the diagnosis of ovarian malignancy. Obstet Gynecol 1994;83(3):434-7.

Brown DL, Doubilet PM, Miller FH, et al. Benign and malignant ovarian masses: selection of the most discriminating gray-scale and Doppler sonographic features. Radiology 1998;208(1):103-10.

Buckshee K, Temsu I, Bhatla N, et al. Pelvic examination, transvaginal ultrasound and transvaginal color Doppler sonography as predictors of ovarian cancer. Int J Gynaecol Obstet 1998;61(1):51-7.

Buist MR, Golding RP, Burger CW, et al. Comparative evaluation of diagnostic methods in ovarian carcinoma with emphasis on CT and MRI. Gynecol Oncol 1994;52(2):191-8.

Buy JN, Ghossain MA, Hugol D, et al. Characterization of adnexal masses: combination of color Doppler and conventional sonography compared with spectral Doppler analysis alone and conventional sonography alone. AJR Am J Roentgenol 1996;166(2):385-93.

Canis M, Mage G, Pouly JL, et al. Laparoscopic diagnosis of adnexal cystic masses: a 12-year experience with long-term follow-up. Obstet Gynecol 1994;83(5 Pt 1):707-12.

Canis M, Mashiach R, Wattiez A, et al. Frozen section in laparoscopic management of macroscopically suspicious ovarian masses. J Am Assoc Gynecol Laparosc 2004;11(3):365-9.

Canis M, Pouly JL, Wattiez A, et al. Laparoscopic management of adnexal masses suspicious at ultrasound. Obstet Gynecol 1997;89(5 Pt 1):679-83.

Carley ME, Klingele CJ, Gebhart JB, et al. Laparoscopy versus laparotomy in the management of benign unilateral adnexal masses. J Am Assoc Gynecol Laparosc 2002;9(3):321-6.

Carter JR, Lau M, Fowler JM, et al. Blood flow characteristics of ovarian tumors: implications for ovarian cancer screening. Am J Obstet Gynecol 1995;172(3):901-7.

Carter PG, Iles RK, Neven P, et al. The measurement of urinary beta core fragment in conjunction with serum CA125 does not aid the differentiation of malignant from benign pelvic masses. Gynecol Oncol 1993;51(3):368-71.

Caruso A, Caforio L, Testa AC, et al. Transvaginal color Doppler ultrasonography in the presurgical characterization of adnexal masses. Gynecol Oncol 1996;63(2):184-91.

Castillo G, Alcazar JL, Jurado M. Natural history of sonographically detected simple unilocular adnexal cysts in asymptomatic postmenopausal women. Gynecol Oncol 2004;92(3):965-9.

Chalas E, Welshinger M, Engellener W, et al. The clinical significance of thrombocytosis in women presenting with a pelvic mass. Am J Obstet Gynecol 1992;166(3):974-7.

Chapron C, Dubuisson JB, Capella-Allouc S. Salpingo-oophorectomy for adnexal masses. Place and results for operative laparoscopy. Eur J Obstet Gynecol Reprod Biol 1997;73(1):43-8.

Chen DX, Schwartz PE, Li FQ. Saliva and serum CA 125 assays for detecting malignant ovarian tumors. Obstet Gynecol 1990;75(4):701-4.

Chen DX, Schwartz PE, Li XG, et al. Evaluation of CA 125 levels in differentiating malignant from benign tumors in patients with pelvic masses. Obstet Gynecol 1988;72(1):23-7.

Chi DS, Abu-Rustum NR, Sonoda Y, et al. Ten-year experience with laparoscopy on a gynecologic oncology service: analysis of risk factors for complications and conversion to laparotomy. Am J Obstet Gynecol 2004;191(4):1138-45.

Childers JM, Nasseri A, Surwit EA. Laparoscopic management of suspicious adnexal masses. Am J Obstet Gynecol 1996;175(6):1451-7; discussion 1457-9.

Chou CY, Chang CH, Yao BL , et al. Color Doppler ultrasonography and serum CA 125 in the differentiation of benign and malignant ovarian tumors. J Clin Ultrasound 1994;22:491-6.

Cohen LS, Escobar PF, Scharm C, et al. Three-dimensional power Doppler ultrasound improves the diagnostic accuracy for ovarian cancer prediction. Gynecol Oncol 2001;82(1):40-8.

Davies AP, Jacobs I, Woolas R, et al. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. Br J Obstet Gynaecol 1993;100(10):927-31.

Deckardt R, Saks M, Graeff H. Comparison of minimally invasive surgery and laparotomy in the treatment of adnexal masses. J Am Assoc Gynecol Laparosc 1994;1(4 Pt 1):333-8.

DePriest PD, Gallion HH, Pavlik EJ, et al. Transvaginal sonography as a screening method for the detection of early ovarian cancer. Gynecol Oncol 1997;65(3):408-14.

DePriest PD, Shenson D, Fried A, et al. A morphology index based on sonographic findings in ovarian cancer. Gynecol Oncol 1993;51(1):7-11.

DePriest PD, van Nagell JR Jr, Gallion HH, et al. Ovarian cancer screening in asymptomatic postmenopausal women. Gynecol Oncol 1993;51(2):205-9.

DePriest PD, Varner E, Powell J. The efficacy of a sonographic morphology index in identifying ovarian cancer: a multi-institutional investigation. Gynecol Oncol 1994;55:174-8.

Dottino PR, Levine DA, Ripley DL, et al. Laparoscopic management of adnexal masses in premenopausal and postmenopausal women. Obstet Gynecol 1999;93(2):223-8.

Dowd JR, Quinn MA, Rome R, et al. Women with a pelvic mass--when to perform an ultrasound. Aust N Z J Obstet Gynaecol 1993;33(4):404-7.

Einhorn N, Bast RC Jr, Knapp RC, et al. Preoperative evaluation of serum CA 125 levels in patients with primary epithelial ovarian cancer. Obstet Gynecol 1986;67(3):414-6.

Ekerhovd E, Wienerroith H, Staudach A, et al. Preoperative assessment of unilocular adnexal cysts by transvaginal ultrasonography: a comparison between ultrasonographic morphologic imaging and histopathologic diagnosis. Am J Obstet Gynecol. 2001;184(2):48-54.

Fanfani F, Fagotti A, Ercoli A, et al. A prospective randomized study of laparoscopy and minilaparotomy in the management of benign adnexal masses. Hum Reprod 2004;19(10):2367-71.

Fenchel S, Grab D, Nuessle K, et al. Asymptomatic adnexal masses: correlation of FDG PET and histopathologic findings. Radiology 2002;223(3):780-8.

Ferdeghini M, Gadducci A, Prontera C, et al. Serum soluble interleukin-2 receptor assay in epithelial ovarian cancer. Tumour Biol 1993;14(5):303-9.

Ferrazzi E, Zanetta G, Dordoni D, et al. Transvaginal ultrasonographic characterization of ovarian masses: comparison of five scoring systems in a multicenter study.[erratum appears in Ultrasound Obstet Gynecol 1998 Jan;11(1):v Note: Lissoni G[corrected to Lissoni AA]]. Ultrasound Obstet Gynecol 1997;10(3):192-7.

Finkler NJ, Benacerraf B, Lavin PT, et al. Comparison of serum CA 125, clinical impression, and ultrasound in the preoperative evaluation of ovarian masses. Obstet Gynecol 1988;72(4):659-64.

Fleischer AC, Cullinan JA, Jones HW 3rd, et al. Correlation of histomorphology of ovarian masses with color Doppler sonography. Ultrasound Med Biol 1996;22(5):555-9.

Fleischer AC, Rodgers WH, Kepple DM, et al. Color Doppler sonography of benign and malignant ovarian masses. Radiographics 1992;12(5):879-85.

Franchi M, Beretta P, Ghezzi F, et al. Diagnosis of pelvic masses with transabdominal color Doppler, CA 125 and ultrasonography. Acta Obstet Gynecol Scand 1995;74(9):734-9.

Gadducci A, Baicchi U, Marrai R, et al. Preoperative evaluation of D-dimer and CA 125 levels in differentiating benign from malignant ovarian masses. Gynecol Oncol 1996;60(2):197-202.

Gadducci A, Capriello P, Bartolini T, et al. The association of ultrasonography and CA-125 test in the preoperative evaluation of ovarian carcinoma. Eur J Gynaecol Oncol 1988;9(5):373-6.

Gadducci A, Ferdeghini M, Prontera C, et al. The concomitant determination of different tumor markers in patients with epithelial ovarian cancer and benign ovarian masses: relevance for differential diagnosis. Gynecol Oncol 1992;44(2):147-54.

Gadducci A, Ferdeghini M, Rispoli G, et al. Comparison of tumor-associated trypsin inhibitor (TATI) with CA125 as a marker for diagnosis and monitoring of epithelial ovarian cancer. Scand J Clin Lab Invest Suppl 1991;207(19-24).

Goldstein SR, Subramanyam B, Snyder JR, et al. The postmenopausal cystic adnexal mass: the potential role of ultrasound in conservative management. Obstet Gynecol 1989;73(1):8-10.

Grab D, Flock F, Stohr I, et al. Classification of asymptomatic adnexal masses by ultrasound, magnetic resonance imaging, and positron emission tomography. Gynecol Oncol 2000;77(3):454-9.

Granberg S, Norstrom A, Wikland M. Tumors in the lower pelvis as imaged by vaginal sonography. Gynecol Oncol 1990;37(2):224-9.

Granberg S, Norstrom A, Wikland M. Comparison ofendovaginal ultrasound and cytological evaluation of cystic ovarian tumors. J Ultrasound Med 1991;10:9-14.

Grover SR, Quinn MA. Is there any value in bimanual pelvic examination as a screening test. Med J Aust 1995;162(8):408-10.

Guerriero S, Ajossa S, Garau N, et al. Ultrasonography and color Doppler-based triage for adnexal masses to provide the most appropriate surgical approach. Am J Obstet Gynecol 2005;192(2):401-6.

Guerriero S, Ajossa S, Risalvato A, et al. Diagnosis of adnexal malignancies by using color Doppler energy imaging as a secondary test in persistent masses. Ultrasound Obstet Gynecol 1998;11(4):277-82.

Guerriero S, Alcazar JL, Coccia ME, et al. Complex pelvic mass as a target of evaluation of vessel distribution by color Doppler sonography for the diagnosis of adnexal malignancies: results of a multicenter European study. J Ultrasound Med 2002;21(10):1105-11.

Hata K, Hata T, Kitao M. Intratumoral peak systolic velocity as a new possible predictor for detection of adnexal malignancy. Am J Obstet Gynecol 1995;172:1496-500.

Hata K, Hata T, Manabe A, et al. A critical evaluation of transvaginal Doppler studies, transvaginal sonography, magnetic resonance imaging, and CA 125 in detecting ovarian cancer. Obstet Gynecol 1992;80(6):922-6.

Havrilesky LJ, Peterson BL, Dryden DK, et al. Predictors of clinical outcomes in the laparoscopic management of adnexal masses. Obstet Gynecol 2003;102(2):243-51.

Herrmann UJ Jr, Locher GW, Goldhirsch A. Sonographic patterns of ovarian tumors: prediction of malignancy. Obstet Gynecol 1987;69(5):777-81.

Hidlebaugh DA, Vulgaropulos S, Orr RK. Treating adnexal masses. Operative laparoscopy vs. laparotomy. J Reprod Med 1997;42(9):551-8.

Hillaby K, Aslam N, Salim R, et al. The value of detection of normal ovarian tissue (the 'ovarian crescent sign') in the differential diagnosis of adnexal masses. Ultrasound Obstet Gynecol 2004;23(1):63-7.

Hogdall EV, Hogdall CK, Tingulstad S, et al. Predictive values of serum tumour markers tetranectin, OVX1, CASA and CA125 in patients with a pelvic mass. Int J Cancer 2000;89(6):519-23.

Hricak H, Chen M, Coakley FV, et al. Complex adnexal masses: detection and characterization with MR imaging--multivariate analysis. Radiology 2000;214(1):39-46.

Huber S, Medl M, Baumann L, et al. Value of ultrasound and magnetic resonance imaging in the preoperative evaluation of suspected ovarian masses. Anticancer Res 2002;22(4):2501-7.

Hurteau JA, Woolas RP, Jacobs IJ, et al. Soluble interleukin-2 receptor alpha is elevated in sera of patients with benign ovarian neoplasms and epithelial ovarian cancer. Cancer 1995;76(9):1615-20.

Inoue M, Fujita M, Nakazawa A, et al. Sialyl-Tn, sialyl-Lewis Xi, CA 19-9, CA 125, carcinoembryonic antigen, and tissue polypeptide antigen in differentiating ovarian cancer from benign tumors. Obstet Gynecol 1992;79(3):434-40.

Itakura T, Kikkawa F, Kajiyama H, et al. Doppler flow and arterial location in ovarian tumors. Int J Gynaecol Obstet 2003;83(3):277-83.

Jacobs I, Oram D, Fairbanks J, et al. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. Br J Obstet Gynaecol 1990;97(10):922-9.

Jacobs I, Stabile I, Bridges J, et al. Multimodal approach to screening for ovarian cancer. Lancet 1988;1(8580):268-71.

Jain KA. Prospective evaluation of adnexal masses with endovaginal gray-scale and duplex and color Doppler US: correlation with pathologic findings. Radiology 1994;191(1):63-7.

Jain KA, Friedman DL, Pettinger TW, et al. Adnexal masses: comparison of specificity of endovaginal US and pelvic MR imaging. Radiology 1993;186(3):697-704.

Juhasz B, Kurjak A, Lampe L, et al. Tissue characterization by transvaginal colour doppler for the evaluation of gynaecological tumours. Acta Med Hung 1990;47:149-56.

Kawahara K, Yoshida Y, Kurokawa T, et al. Evaluation of positron emission tomography with tracer 18-fluorodeoxyglucose in addition to magnetic resonance imaging in the diagnosis of ovarian cancer in selected women after ultrasonography. J Comput Assist Tomogr 2004;28(4):505-16.

Kawai M, Kikkawa F, Ishikawa H. Differential diagnosis of ovarian tumors by transvaginal color -pulse Doppler sonography. Gynecol Oncol 1994;54:209-14.

Komatsu T, Konishi I, Mandai M, et al. Adnexal masses: transvaginal US and gadolinium-enhanced MR imaging assessment of intratumoral structure. Radiology 1996;198(1):109-15.

Kurjak A, Kupesic S. Three dimensional ultrasound and power doppler in assessment of uterine and ovarian angiogenesis: a prospective study. Croat Med J 1999;40(3):413-20.

Kurjak A, Kupesic S, Sparac V, et al. Three-dimensional ultrasonographic and power Doppler characterization of ovarian lesions. Ultrasound Obstet Gynecol 2000;16(4):365-71.

Kurjak A, Predanic M. New scoring system for prediction of ovarian malignancy based on transvaginal color Doppler sonography. J Ultrasound Med 1992;11(12):631-8.

Kurjak A, Schulman H, Sosic A, et al. Transvaginal ultrasound, color flow, and Doppler waveform of the postmenopausal adnexal mass. Obstet Gynecol 1992;80(6):917-21.

Kurjak A, Shalan H, Kupesic S, et al. An attempt to screen asymptomatic women for ovarian and endometrial cancer with transvaginal color and pulsed Doppler sonography. J Ultrasound Med 1994;13(4):295-301.

Kurjak A, Zalud I, Alfirevic Z. Evaluation of adnexal masses with transvaginal color ultrasound. J Ultrasound Med 1991;10(6):295-7.

Kurtz AB, Tsimikas JV, Tempany CM, et al. Diagnosis and staging of ovarian cancer: comparative values of Doppler and conventional US, CT, and MR imaging correlated with surgery and histopathologic analysis--report of the Radiology Diagnostic Oncology Group. Radiology 1999;212(1):19-27.

Leeners B, Schild RL, Funk A, et al. Colour Doppler sonography improves the pre-operative diagnosis of ovarian tumours made using conventional transvaginal sonography. Eur J Obstet Gynecol Reprod Biol 1996;64(1):79-85.

Lerner JP, Timor-Tritsch IE, Federman A, et al. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. Am J Obstet Gynecol 1994;170(1 Pt 1):81-5.

Levine D, Gosink BB, Wolf SI, et al. Simple adnexal cysts: the natural history in postmenopausal women.[see comment]. Radiology 1992;184(3):653-9.

Lin JY, Angel C, DuBeshter B, et al. Diagnoses after laparotomy for a mass in the pelvic area in women. Surg Gynecol Obstet 1993;176(4):333-8.

Lin P, Wu C, Lee C, et al. Color Doppler Ultrasound in the Assessment of Ovarian Neoplasms. J Med Ultrasound 1993;1(4):172-6.

Lok IH, Sahota DS, Rogers MS, et al. Complications of laparoscopic surgery for benign ovarian cysts. J Am Assoc Gynecol Laparosc 2000;7(4):529-34.

Lu C, Van Gestel T, Suykens JA, et al. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. Artif Intell Med 2003;28(3):281-306.

Luxman D, Bergman A, Sagi J, et al. The postmenopausal adnexal mass: correlation between ultrasonic and pathologic findings. Obstet Gynecol 1991;77(5):726-8.

Ma S, Shen K, Lang J. A risk of malignancy index in preoperative diagnosis of ovarian cancer. Chin Med J 2003;116(3):396-9.

Maggino T, Gadducci A, D'Addario V, et al. Prospective multicenter study on CA 125 in postmenopausal pelvic masses. Gynecol Oncol 1994;54(2):117-23.

Malkasian GD Jr, Knapp RC, Lavin PT, et al. Preoperative evaluation of serum CA 125 levels in premenopausal and postmenopausal patients with pelvic masses: discrimination of benign from malignant disease. Am J Obstet Gynecol 1988;159(2):341-6.

Maly Z, Riss P, Deutinger J. Localization of blood vessels and qualitative assessment of blood flow in ovarian tumors. Obstet Gynecol 1995;85(1):33-6.

Mancuso A, De Vivo A, Triolo O, et al. The role of transvaginal ultrasonography and serum CA 125 assay combined with age and hormonal state in the differential diagnosis of pelvic masses. Eur J Gynaecol Oncol 2004;25(2):207-10.

Manjunath AP, Pratapkumar, Sujatha K, et al. Comparison of three risk of malignancy indices in evaluation of pelvic masses. Gynecol Oncol 2001;81(2):225-9.

Mann WJ, Reich H. Laparoscopic adnexectomy in postmenopausal women. J Reprod Med 1992;37(3):254-6.

Marana R, Muzii L, Catalano GF, et al. Laparoscopic excision of adnexal masses. J Am Assoc Gynecol Laparosc 2004;11(2):162-6.

Marchetti M, Zambon A, Lamaina V, et al. Ultrasound as a possible screening method in ovarian cancer.[erratum appears in Eur J Gynaecol Oncol. 2002;23(3):172 Note: Marchioro S [corrected to Marchiori S]]. Eur J Gynaecol Oncol 2002;23(2):123-6.

Marret H, Sauget S, Giraudeau B, et al. Contrast-enhanced sonography helps in discrimination of benign from malignant adnexal masses. J Ultrasound Med. 2004;23(12):1629-39; quiz 1641-42.

Matthes AC, Moreira de Andrade JM , Bighetti S. Selection of criteria for the treatment of ovarian cysts on the bases of ultrasound and cytology. Gynecol Obstet Invest 1996;42:244-8.

McIntosh MW, Drescher C, Karlan B, et al. Combining CA125 and SMR serum markers for diagnosis and early detection of ovarian cancer. Gynecol Oncol 2004;95:9-15.

Medl M, Kulenkampff KJ, Stiskal M, et al. Magnetic resonance imaging in the preoperative evaluation of suspected ovarian masses. Anticancer Res 1995;15(3):1123-5.

Menon U, Talaat A, Rosenthal AN, et al. Performance of ultrasound as a second line test to serum CA125 in ovarian cancer screening. BJOG 2000;107(2):165-9.

Merce LT, Caballero RA, Barco MJ, et al. B-mode, utero-ovarian and intratumoural transvaginal colour Doppler ultrasonography for differential diagnosis of ovarian tumours. Eur J Obstet Gynecol Reprod Biol 1998;76(1):97-107.

Modesitt SC, Pavlik EJ, Ueland FR, et al. Risk of malignancy in unilocular ovarian cystic tumors less than 10 centimeters in diameter. Obstet Gynecol 2003;102(3):594-9.

Mol BW, Boll D, De Kanter M, et al. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. Gynecol Oncol 2001;80(2):162-7.

Morgante G, la Marca A, Ditto A, et al. Comparison of two malignancy risk indices based on serum CA125, ultrasound score and menopausal status in the diagnosis of ovarian masses. Br J Obstet Gynaecol 1999;106(6):524-7.

O'Connell GJ, Ryan E, Murphy KJ, et al. Predictive value of CA 125 for ovarian carcinoma in patients presenting with pelvic masses. Obstet Gynecol 1987;70(6):930-2.

Obeidat BR, Amarin ZO, Latimer JA, et al. Risk of malignancy index in the preoperative evaluation of pelvic masses. Int J Gynaecol Obstet 2004;85(3):255-8.

Ong S, Duffy T, Murphy J. Transabdominal ultrasound and its correlation with clinical findings in gynaecology. Ir J Med Sci 1996;165(4):268-70.

Onsrud M, Shabana A, Austgulen R. Soluble tumor necrosis factor receptors and CA 125 in serum as markers for epithelial ovarian cancer. Tumour Biol 1996;17(2):90-6.

Padilla LA, Radosevich DM, Milad MP. Accuracy of the pelvic examination in detecting adnexal masses. Obstet Gynecol 2000;96(4):593-8.

Padilla LA, Radosevich DM, Milad MP. Limitations of the pelvic examination for evaluation of the female pelvic organs. Int J Gynaecol Obstet 2005;88(1):84-8.

Padungsutt P, Thirapagawong C, Senapad S, et al. Accuracy of tissue polypeptide specific antigen (TPS) in the diagnosis of ovarian malignancy. Anticancer Res 2000;20(2B):1291-5.

Parker J, Proietto AM. Surgical treatment of adnexal cysts by laparotomy. Int J Gynaecol Obstet 1997;59(3):257-8.

Parker WH, Levine RL, Howard FM, et al. A multicenter study of laparoscopic management of selected cystic adnexal masses in postmenopausal women. J Am Coll Surg 1994;179(6):733-7.

Patsner B, Mann WJ. The value of preoperative serum CA 125 levels in patients with a pelvic mass. Am J Obstet Gynecol 1988;159(4):873-6.

Prompeler HJ, Madjar H, Sauerbrei W. Classification of adnexal tumors by transvaginal color Doppler. Gynecol Oncol 1996;61(3):354-63.

Pyrgiotis E, Salamalekis E, Loghis C, et al. CA 125 in preoperative evaluation of pelvic masses. Eur J Gynaecol Oncol 1993;14(4):279-82.

Rehn M, Lohmann K, Rempen A. Transvaginal ultrasonography of pelvic masses: evaluation of B-mode technique and Doppler ultrasonography. Am J Obstet Gynecol 1996;175(1):97-104.

Reles A, Wein U, Lichtenegger W. Transvaginal color Doppler sonography and conventional sonography in the preoperative assessment of adnexal masses. J Clin Ultrasound 1997;25(5):217-25.

Reuter M, Steffens J. , Schuppler U. Critical evaluation of the specificity of MRI and TVUs for differentiation of malignant from benign adnexal lesions. Eur Radiol 1998;8(39-44).

Roman LD, Muderspach LI, Burnett AF, et al. Carcinoembryonic antigen in women with isolated pelvic masses. Clinical utility? J Reprod Med 1998;43(5):403-7.

Roman LD, Muderspach LI, Stein SM, et al. Pelvic examination, tumor marker level, and gray-scale and Doppler sonography in the prediction of pelvic cancer. Obstet Gynecol 1997;89(4):493-500.

Sadik S, Onoglu AS, Gokdeniz R, et al. Laparoscopic management of selected adnexal masses. J Am Assoc Gynecol Laparosc 1999;6(3):313-6.

Salem S, White LM, Lai J. Doppler sonography of adnexal masses: the predictive value of the pulsatility index in benign and malignant disease. AJR Am J Roentgenol 1994;163(5):1147-50.

Sassone AM, Timor-Tritsch IE, Artner A, et al. Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy. Obstet Gynecol 1991;78(1):70-6.

Sawicki W, Spiewankiewicz B, Cendrowski K, et al. Preoperative discrimination between malignant and benign adnexal masses with transvaginal ultrasonography and colour blood flow imaging. Eur J Gynaecol Oncol 2001;22(2):137-42.

Schapira MM, Matchar DB, Young MJ. The effectiveness of ovarian cancer screening. A decision analysis model. Ann Intern Med 1993;118(11):838-43.

Schelling M, Braun M, Kuhn W, et al. Combined transvaginal B-mode and color Doppler sonography for differential diagnosis of ovarian tumors: results of a multivariate logistic regression analysis. Gynecol Oncol 2000;77(1):78-86.

Schincaglia P, Brondelli L, Cicognani A, et al. A feasibility study of ovarian cancer screening: does fine-needle aspiration improve ultrasound specificity? Tumori 1994;80(3):181-7.

Schneider VL, Schneider A, Reed KL, et al. Comparison of Doppler with two-dimensional sonography and CA 125 for prediction of malignancy of pelvic masses. Obstet Gynecol 1993;81(6):983-8.

Schutter EM, Davelaar EM, van Kamp GJ, et al. The differential diagnostic potential of a panel of tumor markers (CA 125, CA 15-3, and CA 72-4 antigens) in patients with a pelvic mass. Am J Obstet Gynecol 2002;187(2):385-92.

Schutter EM, Kenemans P, Sohn C, et al. Diagnostic value of pelvic examination, ultrasound, and serum CA 125 in postmenopausal women with a pelvic mass. An international multicenter study. Cancer 1994;74(4):1398-406.

Schutter EM, Sohn C, Kristen P, et al. Estimation of probability of malignancy using a logistic model combining physical examination, ultrasound, serum CA 125, and serum CA 72-4 in postmenopausal women with a pelvic mass: an international multicenter study. Gynecol Oncol 1998;69(1):56-63.

Scoutt LM, McCarthy SM, Lange R, et al. MR evaluation of clinically suspected adnexal masses. J Comput Assist Tomogr 1994;18(4):609-18.

Sengoku K, Satoh T, Saitoh S, et al. Evaluation of transvaginal color Doppler sonography, transvaginal sonography and CA 125 for prediction of ovarian malignancy. Int J Gynaecol Obstet 1994;46(1):39-43.

Serur E, Emeney PL, Byrne DW. Laparoscopic management of adnexal masses. J Soc Laparoendosc Surg 2001;5(2):143-51.

Shabana A, Onsrud M. Tissue polypeptide-specific antigen and CA 125 as serum tumor markers in ovarian carcinoma. Tumour Biol 1994;15(6):361-7.

Shalev E, Eliyahu S, Peleg D, et al. Laparoscopic management of adnexal cystic masses in postmenopausal women. Obstet Gynecol 1994;83(4):594-6.

Shen-Gunther J, Mannel RS. Ascites as a predictor of ovarian malignancy. Gynecol Oncol 2002;87(1):77-83.

Siegel BA, Dehdashti F, Mutch DG, et al. Evaluation of 111In-DTPA-folate as a receptor-targeted diagnostic agent for ovarian cancer: initial clinical results. J Nucl Med 2003;44(5):700-7.

Skates SJ, Singer DE. Quantifying the potential benefit of CA 125 screening for ovarian cancer. J Clin Epidemiol 1991;44(4-5):365-80.

Smikle CB, Lunt CC, Hankins GD. Clinical predictors in the evaluation of a pelvic mass. Mil Med 1995;160(5):233-5.

Sohaib SA, Mills TD, Sahdev A, et al. The role of magnetic resonance imaging and ultrasound in patients with adnexal masses. Clin Radiol 2005;60(3):340-8.

Somigliana E, Ragni G, Benedetti F, et al. Does laparoscopic excision of endometriotic ovarian cysts significantly affect ovarian reserve? Insights from IVF cycles. Hum Reprod 2003;18(11):2450-3.

Soper JT, Hunter VJ, Daly L, et al. Preoperative serum tumor-associated antigen levels in women with pelvic masses. Obstet Gynecol 1990;75(2):249-54.

Stein SM, Laifer-Narin S, Johnson MB, et al. Differentiation of benign and malignant adnexal masses: relative value of gray-scale, color Doppler, and spectral Doppler sonography. AJR Am J Roentgenol 1995;164(2):381-6.

Strigini FA, Gadducci A, Del Bravo B, et al. Differential diagnosis of adnexal masses with transvaginal sonography, color flow imaging, and serum CA 125 assay in pre- and postmenopausal women. Gynecol Oncol 1996;61(1):68-72.

Szpurek D, Moszyniki R, Sajdak S. Clinical value of the ultrasound Doppler index in determination of ovarian tumor malignancy. Eur J Gynaecol Oncol 2004;25(4):442-4.

Tailor A, Bourne TH, Campbell S, et al. Results from an ultrasound-based familial ovarian cancer screening clinic: a 10-year observational study. Ultrasound Obstet Gynecol 2003;21(4):378-85.

Tailor A, Jurkovic D, Bourne TH, et al. Sonographic prediction of malignancy in adnexal masses using an artificial neural network. Br J Obstet Gynaecol 1999;106(1):21-30.

Takac I. Analysis of blood flow in adnexal tumors by using color Doppler imaging and pulsed spectral analysis. Ultrasound Med Biol 1998;24(8):1137-42.

Tangjitgamol S, Jesadapatrakul S, Manusirivithaya S, et al. Accuracy of frozen section in diagnosis of ovarian mass. Int J Gynecol Cancer 2004;14(2):212-9.

Tanir HM, Ozalp S, Yalcin OT, et al. Preoperative serum vascular endothelial growth factor (VEGF) in ovarian masses. Eur J Gynaecol Oncol 2003;24(3-4):271-4.

Tarik A, Fehmi C. Complications of gynaecological laparoscopy--a retrospective analysis of 3572 cases from a single institute. J Obstet Gynaecol 2004;24(7):813-6.

Tay SK, Chua EK. Correlation of serum, urinary and salivary CA 125 levels in patients with adnexal masses. Ann Acad Med Singapore 1994;23(3):311-4.

Tekay A, Jouppila P. Validity of pulsatility and resistance indices in classification of adnexal tumors with transvaginal color doppler ultrasound. Ultrasound Obstet Gynecol 1992;2:338-44.

Tengs TO, Winer EP, Paddock S, et al. Testing for the BRCA1 and BRCA2 breast-ovarian cancer susceptibility genes: a decision analysis. Med Decis Making 1998;18(4):365-75.

Tepper R, Lerner-Geva L, Altaras MM, et al. Transvaginal color flow imaging in the diagnosis of ovarian tumors. J Ultrasound Med 1995;14(10):731-4.

Tian J, Zhang J, Jiao L, et al. A prospective study of Tc-99m MIBI in the differential diagnosis of pelvic masses in female patients. Clin Nucl Med 2000;25(8):614-8.

Timmerman D, Bourne TH, Tailor A, et al. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model. Am J Obstet Gynecol 1999;181(1):57-65.

Timmerman D, Verrelst H, Bourne TH, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound Obstet Gynecol 1999;13(1):17-25.

Timor-Tritsch LE, Lerner J, Monteagudo A, et al. Transvaginal ultrasonographic characterization of ovarian masses by means of color flow-directed Doppler measurements and a morphologic scoring system. Am J Obstet Gynecol 1993;168:909-13.

Tingulstad S, Hagen B, Skjeldestad FE, et al. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. Obstet Gynecol 1999;93(3):448-52.

Tingulstad S, Hagen B, Skjeldestad FE, et al. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. Br J Obstet Gynaecol 1996;103(8):826-31.

Torres JC, Derchain SF, Faundes A, et al. Risk-of-malignancy index in preoperative evaluation of clinically restricted ovarian cancer. Rev Paul Med 2002;120(3):72-6.

Troiano RN, Quedens-Case C, Taylor KJ. Correlation of findings on transvaginal sonography with serum CA 125 levels. AJR Am J Roentgenol 1997;168(6):1587-90.

Twickler DM, Forte TB, Santos-Ramos R, et al. The Ovarian Tumor Index predicts risk for malignancy. Cancer 1999;86(11):2280-90.

Urban N, Drescher C, Etzioni R, et al. Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening. Control Clin Trials 1997;18(3):251-70.

Valentin L. Gray scale sonography, subjective evaluation of the color Doppler image and measurement of blood flow velocity for distinguishing benign and malignant tumors of suspected adnexal origin. Eur J Obstet Gynecol Reprod Biol 1997;72(1):63-72.

Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. Ultrasound Obstet Gynecol 1999;14(5):338-47.

Valentin L. Comparison of Lerner score, Doppler ultrasound examination, and their combination for discrimination between benign and malignant adnexal masses. Ultrasound Obstet Gynecol 2000;15(2):143-7.

Valentin L, Akrawi D. The natural history of adnexal cysts incidentally detected at transvaginal ultrasound examination in postmenopausal women. Ultrasound Obstet Gynecol 2002;20(2):174-80.

Valentin L, Hagen B, Tingulstad S, et al. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. Ultrasound Obstet Gynecol 2001;18(4):357-65.

van Herendael B, Beretta P, Slangen T, et al. Management of adnexal masses by operative laparoscopy. J Am Assoc Gynecol Laparosc 1995;2(3):273-7.

van Nagell JR Jr, DePriest PD, Reedy MB, et al. The efficacy of transvaginal sonographic screening in asymptomatic women at risk for ovarian cancer. Gynecol Oncol 2000;77(3):350-6.

Vasilev SA, Schlaerth JB, Campeau J, et al. Serum CA 125 levels in preoperative evaluation of pelvic masses. Obstet Gynecol 1988;71(5):751-6.

Vuento MH, Pirhonen JP, Makinen JI, et al. Evaluation of ovarian findings in asymptomatic postmenopausal women with color Doppler ultrasound. Cancer 1995;76(7):1214-8.

Wakahara F, Kikkawa F, Nawa A, et al. Diagnostic efficacy of tumor markers, sonography, and intraoperative frozen section for ovarian tumors. Gynecol Obstet Invest 2001;52(3):147-52.

Weiner Z, Thaler I, Beck D, et al. Differentiating malignant from benign ovarian tumors with transvaginal color flow imaging. Obstet Gynecol 1992;79(2):159-62.

Woolas RP, Conaway MR, Xu F, et al. Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. Gynecol Oncol 1995;59(1):111-6.

Wu CC, Lee CN, Chen T, et al. Factors contributing to the accuracy in diagnosing ovarian malignancy by color Doppler ultrasound. Obstet Gynecol 1994;84:605-8.

Yamashita Y, Hatanaka Y, Torashima M, et al. Characterization of sonographically indeterminate ovarian tumors with MR imaging. A logistic regression analysis. Acta Radiol 1997;38(4 Pt 1):572-7.

Yamashita Y, Torashima M, Hatanaka Y, et al. Adnexal masses: accuracy of characterization with transvaginal US and precontrast and postcontrast MR imaging. Radiology 1995;194(2):557-65.

Yuen PM, Yu KM, Yip SK, et al. A randomized prospective study of laparoscopy and laparotomy in the management of benign ovarian masses. Am J Obstet Gynecol 1997;177(1):109-14.

Zanetta G, Vergani P, Lissoni A. Color Doppler ultrasound in the preoperative assessment of adnexal masses. Acta Obstet Gynecol Scand 1994;73:637-41.

Zhang Z, Barnhill SD, Zhang H, et al. Combination of multiple serum markers using an artificial neural network to improve specificity in discriminating malignant from benign pelvic masses. Gynecol Oncol 1999;73(1):56-61.

Zimmer Y, Tepper R, Akselrod S. An automatic approach for morphological analysis and malignancy evaluation of ovarian masses using B-scans. Ultrasound Med Biol 2003;29(11):1561-70.

List of Acronyms/Abbreviations

2D	Two-dimensional
3D	Three-dimensional
ACOG	American College of Obstetricians and Gynecologists
ACR	American College of Radiology
AFP	Alpha-fetoprotein
AHRQ	Agency for Healthcare Research and Quality
AUC	Area under the curve
CA-125	Cancer antigen 125
CDC	Centers for Disease Control and Prevention
CEA	Carcinoembryonic antigen
CI	Confidence interval
CMS	Centers for Medicaid and Medicare Services
CT	Computed tomography
FDG	18-Fluorodeoxyglucose
FIGO	International Federation of Gynecology and Obstetrics
FNA	Fine needle aspiration
hCG	Human chorionic gonadotropin
ICD-9	*International Classification of Diseases, Ninth Revision*
LDH	Lactate dehydrogenase
LMP	Low malignant potential
MeSH	Medical Subject Heading
MRI	Magnetic resonance imaging
NIS	Nationwide Inpatient Sample
NPV	Negative predictive value
PET	Positron emission tomography
PI	Pulsatility index
PPV	Positive predictive value
RI	Resistance index
RMI	Risk of Malignancy Index
ROC	Receiver operating characteristic
SEER	Surveillance, Epidemiology, and End Results
SGO	Society of Gynecologic Oncologists
TAG-72	Tumor-associated glycoprotein 72
TVUS	Transvaginal ultrasound

APPENDIXES

to

“Management of Adnexal Mass”

Prepared by the Duke Evidence-based Practice Center
(Contract #290-02-0025)

Appendix A: Exact Search Strings

Search Strategy 1: pelvic exam performance
(developed and run by McCrory and Myers on September 10, 2004)

Database: Ovid MEDLINE(R) <1966 to September Week 1 2004>
Search Strategy:
---

1 pelvic exam.mp.(53)
2 (bimanual adj pelvic).mp. [mp=title, original title, abstract, name of substance, mesh subject heading] (25)
3 (physical exam and pelvis).mp.(7)
4 "diagnostic techniques, obstetrical and gynecological"/ or culdoscopy/ or laparoscopy/ or physical examination/ (45383)
5 physical examination/ (18265)
6 Ovarian Cysts/ or Ovarian Neoplasms/ or Genital Neoplasms, Female/ or Adnexal Diseases/ or adnexal mass.mp. (48599)
7 exp Ovarian Cysts/ or exp Ovarian Neoplasms/ or Genital Neoplasms, Female/ or Adnexal Diseases/ or adnexal mass.mp. (53879)
8 exp fallopian tube diseases/ (4449)
9 5 and (7 or 8) (124)
10 (or/1-3) and (or/7-8) (18)
11 9 and 10 (5)
12 "diagnostic techniques, obstetrical and gynecological"/ and (or/7-8) (8)
13 culdoscopy/ and (or/7-8) (52)
14 or/1-3,9-10 (204)
15 limit 14 to (human and english language and yr=1980 - 2004) (147)
16 from 15 keep 1-147 (147)

****

Search Strategy 2: test performance
Developed and run by McCrory on September 28, 2004

Database: Ovid MEDLINE(R) <1966 to September Week 3 2004>
Search Strategy:
---

1 (vagin$ adj ultraso$).mp. [mp=title, original title, abstract, name of substance, mesh subject heading] (1391)
2 (adnex$ adj2 mas$).mp. (873)
3 (pelvi$ adj mas$).mp. (1537)
4 (ovar$ adj mas$).mp. (1479)
5 or/2-4 (3696)
6 "sensitivity and specificity"/ (121128)
7 6 and 1 (132)

8 6 and 5 (316)
9 7 or 8 (431)
10 limit 9 to (human and english language) (387)
11 from 10 keep 1-387 (387)
12 (ovar$ adj tumo$).mp. (11435)
13 12 and 6 (405)
14 ROC Curve/ (7282)
15 13 and 14 (27)
16 from 15 keep 4,7,9,15,19-20,22-23,27 (9)
17 from 15 keep 22-23,27 (3)
18 16 not 11 (4)
19 11 or 18 (391)
20 limit 19 to yr=1980 - 2004 (391)
21 from 20 keep 1-391 (391)

****

Search Strategy 3: predictive models
(strategy developed and run by McCrory on September 29, 2004)

Database: Ovid MEDLINE(R) <1966 to September Week 3 2004>
Search Strategy:
---

1 (vagin$ adj ultraso$).mp. [mp=title, original title, abstract, name of substance, mesh subject heading] (1391)
2 (adnex$ adj2 mas$).mp. (873)
3 (pelvi$ adj mas$).mp. (1537)
4 (ovar$ adj mas$).mp. (1479)
5 or/2-4 (3696)
6 "sensitivity and specificity"/ (121128)
7 6 and 1 (132)
8 6 and 5 (316)
9 7 or 8 (431)
10 limit 9 to (human and english language) (387)
11 predictive value of tests/ (56850)
12 Risk Assessment/ (47548)
13 roc curve/ (7282)
14 "Multivariate Analysis"/ (31714)
15 or/11-14 (136223)
16 15 and 5 (260)
17 16 not 9 (142)
18 limit 17 to (human and english language) (131)
19 from 18 keep 1-131 (131)

****

Appendix B: List of Excluded Studies

All excluded studies listed below were reviewed in their full text version. Following each reference, in italics, is the reason(s) for exclusion and the Question (Q) for which the article was considered. If no Q is indicated, then the article was excluded a priori from the study for the reason given. An article can be considered (and therefore excluded) for more than one question, and all questions for which the article was excluded are identified. Reasons for exclusion signify only the usefulness of the articles for this study and are not intended as criticisms of the articles.

For reference, the questions are:

Question 1: What is the prevalence of various tumor types among women with an adnexal mass, stratified by cancer status (malignant vs. benign), age, menopausal status, and size of tumor?

Question 2: What are the sensitivity, specificity, and reliability of the bimanual examination?

Question 3: Among women with a palpable adnexal mass on exam or a mass identified by ultrasound/imaging, what is the sensitivity/specificity of various evaluation modalities including ultrasound (transvaginal ultrasound, transabdominal ultrasound, color Doppler, 2D vs. 3D ultrasound, CT scan, MRI scan, and CA-125 levels) for diagnosing malignant masses?

Question 4: What is the accuracy of explicit scoring systems which incorporate various combinations of imaging findings, patient risk factors, and/or CA-125 levels for detecting malignancy? Have these scoring systems been applied to a population of women before laparoscopy?

Question 5: Among women with suspected benign lesions on initial investigation, what are the sensitivity and specificity of monitoring with periodic CA-125 and/or interval ultrasound examinations for detecting malignant masses? How does the interval of testing/definition of change affect sensitivity and predictive value?

Question 6: Among women with adnexal masses, what are the morbidity and mortality from diagnostic surgery (laparoscopy or laparotomy)? At what point does the risk of laparoscopy outweigh the risk of detecting malignancy?

Question 7: What are the estimated trade-offs resulting from various strategies for evaluation of the adnexal mass?

Abu-Rustum NR, Rhee EH, Chi DS, et al. Subcutaneous tumor implantation after laparoscopic procedures in women with malignant disease.[see comment]. Obstet Gynecol 2004;103(3):480-7. *Exclude no mass.*

Adonakis GL, Paraskevaidis E, Tsiga S, et al. A combined approach for the early detection of ovarian cancer in asymptomatic women. Eur J Obstet Gynecol Reprod Biol 1996;65(2):221-5. *Exclude Q5-wrong pt population.*

Alcazar JL, Jurado M. Using a logistic model to predict malignancy of adnexal masses based on menopausal status, ultrasound morphology, and color Doppler findings. Gynecol Oncol 1998;69(2):146-50. *Exclude Q3-unable to construct 2x2.*

Alcazar JL, Jurado M. Prospective evaluation of a logistic model based on sonographic morphologic and color Doppler findings developed to predict adnexal malignancy. J Ultrasound Med 1999;18(12):837-42. *Exclude Q3-unable to construct 2x2.*

Alcazar JL, Laparte C, Jurado M, et al. The role of transvaginal ultrasonography combined with color velocity imaging and pulsed Doppler in the diagnosis of endometrioma. Fertil Steril 1997;67(3):487-91. *Exclude Q1-sample size .*

Alcazar JL, Ruiz-Perez ML, Errasti T. Transvaginal color Doppler sonography in adnexal masses: which parameter performs best? Ultrasound Obstet Gynecol 1996;8(2):114-9. *Exclude Q3-unable to construct 2x2.*

Alexander-Sefre F, Menon U, Jacobs IJ. Ovarian cancer screening. Hosp Med 2002;63(4):210-3. *Exclude review.*

Ali N, Jan H, Van Trappen P, et al. Radioimmunoscintigraphy with Tc-99m-labelled SM3 in differentiating malignant from benign adnexal masses. BJOG 2003;110(5):508-14. *Exclude Q3-experimental or non-standard test.*

Alvarez RD, Kilgore LC, Partridge EE, et al. Staging ovarian cancer diagnosed during laparoscopy: accuracy rather than immediacy. South Med J 1993;86(11):1256-8. *Exclude review.*

Alvarez-Sanchez F, Brache V, de Oca VM, et al. Prevalence of enlarged ovarian follicles among users of levonorgestrel subdermal contraceptive implants (Norplant). Am J Obstet Gynecol 2000;182(3):535-9. *Exclude Q3-no histol. dx.*

American College of Obstetricians and Gynecologists. ACOG Committee Opinion: number 280, December 2002. The role of the generalist obstetrician-gynecologist in the early detection of ovarian cancer. Obstet Gynecol 2002;100(6):1413-6. *Exclude review.*

Anderiesz C, Quinn MA. Screening for ovarian cancer. Med J Aust 2003;178(12):655-6. *Exclude review.*

Andersen WA, Nichols GE, Avery SR, et al. Cytologic diagnosis of ovarian tumors: factors influencing accuracy in previously undiagnosed cases. Am J Obstet Gynecol 1995;173(2):457-63; discussion 463-4. *Exclude Q3-wrong test.*

Anderson MM, Irwin CE Jr, Snyder DL. Abnormal vaginal bleeding in adolescents. Pediatr Ann 1986;15(10):697-701. *Exclude Q1-no histol. dx.*

Andolf E, Jorgensen C, Astedt B. Ultrasound examination for detection of ovarian carcinoma in risk groups. Obstet Gynecol 1990;75(1):106-9. *Exclude Q7-not descrip of sim model.*

Angeid-Backman E, Coleman BG, Arger PH, et al. Comparison of resistive index versus pulsatility index in assessing the benign etiology of adnexal masses. Clin Imaging 1998;22(4):284-91. *Exclude no mass.*

Aslam N, Tailor A, Lawton F, et al. Prospective evaluation of three different models for the pre-operative diagnosis of ovarian cancer. BJOG 2000;107(11):1347-53. *Exclude Q3-inconsistent data.*

Aubel S, Wozney P, Edwards RP. MRI of female uterine and juxta-uterine masses: clinical application in 25 patients. Magn Reson Imaging 1991;9(4):485-91.*Exclude Q3-sample size.*

Bandera CA, Ye B, Mok SC. New technologies for the identification of markers for early detection of ovarian cancer. Curr Opin Obstet Gynecol 2003;15(1):51-5. Exclude review.

Baron AT, Cora EM, Lafky JM, et al. Soluble epidermal growth factor receptor (sEGFR/sErbB1) as a potential risk, screening, and diagnostic serum biomarker of epithelial ovarian cancer. Cancer Epidemiol Biomarkers Prev 2003;12(2):103-13. *Exclude no mass.*

Bast RC Jr, Feeney M, Lazarus H, et al. Reactivity of a monoclonal antibody with human ovarian carcinoma. J Clin Invest 1981;68(5):1331-7. *Exclude no mass*.

Bast RC Jr, Knauf S, Epenetos A, et al. Coordinate elevation of serum markers in ovarian cancer but not in benign disease. Cancer 1991;68(8):1758-63. *Exclude no mass*.

Bast RC Jr, Urban N, Shridhar V, et al. Early detection of ovarian cancer: promise and reality. Cancer Treat Res 2002;107:61-97. *Exclude review.*

Bell R, Petticrew M, Sheldon T. The performance of screening tests for ovarian cancer: results of a systematic review. Br J Obstet Gynaecol 1998;105(11):1136-47. *Exclude no mass.*

Benacerraf BR, Finkler NJ, Wojciechowski C, et al. Sonographic accuracy in the diagnosis of ovarian masses. J Reprod Med 1990;35(5):491-5.*Exclude Q3-distguish malignant versus nonmalignant*.

Berlanda N, Ferrari MM, Mezzopane R, et al. Impact of a multiparameter, ultrasound-based triage on surgical management of adnexal masses. Ultrasound Obstet Gynecol 2002;20(2):181-5. *Exclude Q6-no M&M data.*

Biran G, Golan A, Sagiv R, et al. Conversion of laparoscopy to laparotomy due to adnexal malignancy. Eur J Gynaecol Oncol 2002;23(2):157-60.*Exclude Q6-no M&M data/Exclude Q4-unable to construct 2x2/ Exclude Q3-unable to construct 2x2.*

Blend MJ, Ostrowski GJ. Recent advances in the detection of ovarian cancer: a review.... J Am Osteopath Assoc 1994;94(4):305-18. *Exclude review.*

Bohm-Velez M, Mendelson E, Bree R, et al. Ovarian cancer screening. American College of Radiology. ACR Appropriateness Criteria. Radiology 2000;215(Suppl):861-71. *Exclude review*.

Bohm-Velez M, Mendelson E, Bree R, et al. Suspected adnexal masses. American College of Radiology. ACR Appropriateness Criteria. Radiology 2000;215(Suppl):931-8. *Exclude review* .

Boll D, Geomini PM, Brolmann HA, et al. The pre-operative assessment of the adnexal mass: the accuracy of clinical estimates versus clinical prediction rules. BJOG 2003;110(5):519-23. *Exclude Q4-partial dupl new data not relevant.*

Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Clin Radiol 2003;58(8):575-80. *Exclude review.*

Bourne TH, Campbell S, Reynolds KM, et al. Screening for early familial ovarian cancer with transvaginal ultrasonography and colour blood flow imaging. BMJ 1993;306(6884):1025-9. *Exclude no mass.*

Bourne TH, Hampson J, Reynolds K, et al. Screening for early ovarian cancer. Br J Hosp Med 1992;48(8):454-9. *Exclude review.*

Brown DL, Frates MC, Laing FC, et al. Ovarian masses: can benign and malignant lesions be differentiated with color and pulsed Doppler US? Radiology 1994;190(1):333-6. *Exclude Q3-sample size.*

Brown DL, Zou KH, Tempany CM, et al. Primary versus secondary ovarian malignancy: imaging findings of adnexal masses in the Radiology Diagnostic Oncology Group Study. Radiology 2001;219(1):213-8. *Exclude no mass* .

Buist MR, Golding RP, Burger CW, et al. Comparative evaluation of diagnostic methods in ovarian carcinoma with emphasis on CT and MRI. Gynecol Oncol 1994;52(2):191-8.*Exclude Q6-no M&M data.*

Buquet RA, Amato AR, Huang GB, et al. Is preoperative selection of patients with cystic adnexal masses essential for laparoscopic treatment?. J Am Assoc Gynecol Laparosc 1999;6(4):477-81.*Exclude Q6-no M&M data.*

Buy JN, Ghossain MA, Mark AS, et al. Focal hyperdense areas in endometriomas: a characteristic finding on CT. AJR Am J Roentgenol 1992;159(4):769-71. *Exclude Q3-distguish malignant versus nonmalignant.*

Buy JN, Ghossain MA, Sciot C, et al. Epithelial tumors of the ovary: CT findings and corrrelation with US. Radiology 1991;178:811-18. *Exclude Q3-unable to construct 2x2.*

Campbell S, Bhan V, Royston P, et al. Transabdominal ultrasound screening for early ovarian cancer. BMJ 1989;299(6712):1363-7. *Exclude no mass.*

Canis M, Bassil S, Wattiez A, et al. Fertility following laparoscopic management of benign adnexal cysts. Hum Reprod 1992;7(4):529-31. *Exclude Q6-review.*

Canis M, Pouly JL, Wattiez A, et al. Laparoscopic management of adnexal masses suspicious at ultrasound. Obstet Gynecol 1997;89(5 Pt 1):679-83. *Exclude Q6-no M&M data.*

Caoili EM, Hertzberg BS, Kliewer MA, et al. Refractory shadowing from pelvic masses on sonography: a useful diagnostic sign for uterine leiomyomas. AJR Am J Roentgenol 2000;174(1):97-101. *Exclude no mass* .

Cappelleri JC, Ioannidis JP, Schmid C. Large trials vs meta-analysis of smaller trials: how do their results compare? JAMA 1996;276:1332-8. *Exclude review.*

Carlson KJ, Skates SJ, Singer DE. Screening for ovarian cancer. Ann Intern Med 1994;121(2):124-32. *Exclude review.*

Carter J. An update on ovarian cancer screening. Aust N Z J Obstet Gynaecol 1994;34(2):169-74.
Exclude no mass.

Carter J, Fowler J, Carson L, et al. How accurate is the pelvic examination as compared to transvaginal sonography? A prospective, comparative study. J Reprod Med 1994;39(1):32-4. *Exclude Q3-distguish malignant versus nonmalignant/ Exclude Q2-unable to construct 2x2.*

Carter J, Saltzman A, Hartenbach E, et al. Flow characteristics in benign and malignant gynecologic tumors using transvaginal color flow Doppler. Obstet Gynecol 1994;83(1):125-30. *Exclude Q3-distguish malignant versus nonmalignant.*

Chadha P, Puri M, Gupta R. A comparative evaluation of clinical examination, pelvic ultrasound and laparoscopy in the diagnosis of pelvic masses. Indian J Med Sci 1994;48(7):158-60. *Exclude Q2-unable to construct 2x2.*

Chalas E, Constantino J, Wickerham L, et al. Benign gynecologic conditions among participants in the breast cancer prevention trial. Am J Obstet and Gynecol 2005;192:1230-9. *Exclude Q1-wrong pt population.*

Cherry C, Vacchiano SA. Ovarian cancer screening and prevention. Semin Oncol Nurs 2002;18(3):167-73. *Exclude no mass* .

Childers JM, Aqua KA, Surwit EA, et al. Abdominal-wall tumor implantation after laparoscopy for malignant conditions. Obstet Gynecol 1994;84(5):765-9. *Exclude no mass* .

Close RJ, Sachs CJ, Dyne PL. Reliability of bimanual pelvic examinations performed in emergency departments. West J Med 2001;175(4):240-4. *Exclude Q2-unable to construct 2x2*.

Cohen L, Fishman DA. Ultrasound and ovarian cancer. Cancer Treat Res 2002;107:119-32. *Exclude review* .

Cooper BC, Ritchie JM, Broghammer CL, et al. Preoperative serum vascular endothelial growth factor levels: significance in ovarian cancer. Clin Cancer Res 2002;8(10):3193-7. *Exclude Q3-experimental or non-standard test.*

Crade M, Yiu-Chiu V, Kincaid K. Color Doppler and ovarian masses: familiarity breeds confidence. Ultrasound Obstet Gynecol 1995;6(5):373-4. *Exclude Q3-no histol. dx* .

Crawford RA, Gore ME, Shepherd JH. Ovarian cancers related to minimal access surgery.[see comment]. Br J Obstet Gynaecol 1995;102(9):726-30. *Exclude no mass.*

Crayford TJ, Campbell S, Bourne TH, et al. Benign ovarian cysts and ovarian cancer: a cohort study with implications for screening. Lancet 2000;355(9209):1060-3. *Exclude Q5-wrong pt population.*

Creasman WT, Soper JT. The undiagnosed adnexal mass after the menopause. Clin Obstet Gynecol 1986;29(2):446-52. *Exclude Q1-study design not case series or cohort.*

Crump C, McIntosh MW, Urban N, et al. Ovarian cancer tumor marker behavior in asymptomatic healthy women: implications for screening. Cancer Epidemiol Biomarkers Prev 2000;9(10):1107-11. *Exclude review.*

Crvenkovic G, Karlan BY, Platt LD. Current role of ultrasound in ovarian cancer screening. Clin Obstet Gynecol 1996;39(1):259-67. *Exclude review.*

de Bruijn HW, van der Zee AG, Aalders JG. The value of cancer antigen 125 (CA 125) during treatment and follow-up of patients with ovarian cancer. Curr Opin Obstet Gynecol 1997;9(1):8-13. *Exclude Q3-distguish malignant versus nonmalignant.*

De Vries SO, Hunink MG, Polak JF. Summary receiver operating characteristic curves as a technique for meta analysis of the diganoistic performance of duplex ultrasonography in peripheral arterial disease. Acad Radiol 1996;3:361-9. *Exclude review.*

Decloedt J, Berteloot P, Vergote I. The feasibility of open laparoscopy in gynecologic-oncologic patients. Gynecol Oncol 1997;66(1):138-40. *Exclude no mass.*

Demirkiran F, Kumbak B, Bese T, et al. Vascular endothelial growth factor in adnexal masses. Int J Gynaecol Obstet 2003;83(1):53-8. *Exclude Q3-experimental or non-standard test.*

Dgani R, Shani A, Elchalal U, et al. The leukocyte adherence inhibition test (LAI) in preoperative diagnosis of epithelial ovarian cancer. Gynecol Oncol 1993;49(3):349-53. *Exclude Q3-wrong test.*

Dietrich M, Osmers RG, Grobe G, et al. Limitations of the evaluation of adnexal masses by its macroscopic aspects, cytology and biopsy. Eur J Obstet Gynecol Reprod Biol 1999;82(1):57-62. *Exclude Q3-wrong test.*

Dixon JG, Bognar BA, Keyserling TC, et al. Teaching women's health skills: confidence, attitudes and practice patterns of academic generalist physician. J Gen Intern Med 2003;18(6):411-8. *Exclude review.*

Dogan MM, Ugur M, Soysal SK, et al. Transvaginal sonographic diagnosis of ovarian endometrioma. Int J Gynaecol Obstet 1966;52(2):145-9. *Exclude no mass* .

Domar AD. Psychological aspects of the pelvic exam: individual needs and physician involvement. Women Health 1985-1986;10(4):75-90. *Exclude no mass* .

Dordoni D, Zaglio S, Zucca S, et al. The role of sonographically guided aspiration in the clinical management of ovarian cysts. J Ultrasound Med 1993;12(1):27-31. *Exclude Q3-wrong test.*

Dorum A, Blom G, Ekerhovd E, et al. Prevalence and histologic diagnosis of adnexal cysts in postmenopausal women: an autopsy study. Am J Obstet Gynecol 2005;192:48-54. *Exclude non U.S.*

Dueholm M, Lundorf E, Hansen ES, et al. Magnetic resonance imaging and transvaginal ultrasonography for the diagnosis of adenomyosis. Fertil Steril 2001;76(3):588-94. *Exclude no mass.*

Ehlen T. Management of low malignant potential tumour of the ovary: a policy statement. SOGC/GOC/SCC Policy and Practice Guideline Committee 2000;(85). *Exclude review.*

Einhorn N. Ovarian cancer. Early diagnosis and screening. Hematol Oncol Clin North Am 1992;6(4):843-50. *Exclude no mass* .

Eisen A, Rebbeck TR, Wood WC, et al. Prophylactic surgery in women with a hereditary predisposition to breast and ovarian cancer. J Clin Oncol 2000;18(9):1980-95. *Exclude review.*

Elg S, Lee RB, Stones C, et al. Evaluation of serum haptoglobin levels in patients with adnexal masses. Mil Med 1989;154(5):234-6. *Exclude Q3-experimental or non-standard test.*

Elit L. Surgical management of an adnexal mass suspicious for malignancy. SOGC Clinical Practice Guidelines 2000;(97). *Exclude review.*

Elwood M. Proteomic patterns in serum and identification of ovarian cancer. Lancet 2002;360(9327):170-1. *Exclude no mass .*

Emery J, Yaphe J, Priest P, et al. Screening for ovarian cancer. Lancet 1999;354(9177):509-10. *Exclude no mass .*

Fadare O, Mariappan MR, Wang S, et al. The histologic subtype of ovarian tumors affects the detection rate by pelvic washings. Cancer 2004;102(3):150-6. *Exclude no mass .*

Fayed ST, Ahmad SM, Kassim SK, et al. The value of CA 125 and CA72-4 in management of patients with epithelial ovarian cancer. Dis Markers 1998;14(3):155-60. *Exclude Q3-no verification test negative.*

Fedele L, Bianchi S, Dorta M, et al. Transvaginal ultrasonography versus hysteroscopy in the diagnosis of uterine submucous myomas. Obstet Gynecol 1991;77(5):745-8. *Exclude no mass .*

Fedele L, Bianchi S, Dorta M, et al. Transvaginal ultrasonography in the differential diagnosis of adenomyoma versus leiomyoma. Am J Obstet Gynecol 1992;167(3):603-6. *Exclude no mass .*

Finkler NJ. Clinical utility of CA 125 in preoperative diagnosis of patients with pelvic masses. Eur J Obstet Gynecol Reprod Biol 1993;49(1-2):105-7. *Exclude Review .*

Finkler NJ, Benacerraf B, Lavin PT, et al. Comparison of serum CA 125, clinical impression, and ultrasound in the preoperative evaluation of ovarian masses. Obstet Gynecol 1988;72(4):659-64. *Exclude Q1-no histol. dx.*

Fishman DA, Cohen L, Blank SV, et al. The role of ultrasound evaluation in the detection of early-stage epithelial ovarian cancer. Am J Obstet Gynecol 2005;192(4):1214-21; discussion 1221-2. *Exclude Q3-unable to construct 2x2.*

Fleischer AC, Cullinan JA, Jones HW 3rd, et al. Correlation of histomorphology of ovarian masses with color Doppler sonography. Ultrasound Med Biol 1996;22(5):555-9. *Exclude Q3-unable to construct 2x2.*

Fleischer AC, Cullinan JA, Kepple DM, et al. Conventional and color Doppler transvaginal sonography of pelvic masses: a comparison of relative histologic specificities. J Ultrasound Med 1993;12(12):705-12. *Exclude Q3-unable to construct 2x2.*

Fleischer AC, Jones HW 3rd. Color Doppler sonography of ovarian masses: the importance of a multiparameter approach. Gynecol Oncol 1993;50(1):1-2. *Exclude review .*

Fleischer AC, McKee MS, Gordon AN, et al. Transvaginal sonography of postmenopausal ovaries with pathologic correlation. J Ultrasound Med 1990;9(11):637-44. *Exclude no mass .*

Fleischer AC, Rodgers WH, Rao BK, et al. Assessment of ovarian tumor vascularity with transvaginal color Doppler sonography. J Ultrasound Med 1991;10(10):563-8. *Exclude Q3-publ duplicate.*

Flynn MK, Niloff JM. Outpatient minilaparotomy for ovarian cysts. J Reprod Med 1999;44(5):399-404. *Exclude Q6-no M&M data.*

Foxall MJ, Barron CR, Houfek JF. Ethnic influences on body awareness, trait anxiety, perceived risk, and breast and gynecologic cancer screening practices. Oncol Nurs Forum 2001;28(4):727-38. *Exclude review.*

Franchi M, Beretta P, Ghezzi F, et al. Diagnosis of pelvic masses with transabdominal color Doppler, CA 125 and ultrasonography. Acta Obstet Gynecol Scand 1995;74(9):734-9. *Exclude Q4-unable to construct 2x2.*

Frederick JL, Paulson RJ, Sauer MV. Routine use of vaginal ultrasonography in the preoperative evaluation of gynecologic patients. An adjunct to resident education. J Reprod Med 1991;36(11):779-82. *Exclude no mass .*

Frenkel Y, Oelsner G, Ben-Baruch G, et al. Major surgical complications of laparoscopy. Eur J Obstet Gynecol Reprod Biol 1981;12(2):107-11. *Exclude Other All Premenopausal.*

Gadducci A, Baicchi U, Marrai R, et al. Pretreatment plasma levels of fibrinopeptide-A (FPA), D-dimer (DD), and von Willebrand factor (vWF) in patients with ovarian carcinoma. Gynecol Oncol 1994;53(3):352-6. *Exclude Q3-wrong test.*

Gadducci A, Marrai R, Baicchi U, et al. Preoperative D-dimer plasma assay is not a predictor of clinical outcome for patients with advanced ovarian cancer. Gynecol Oncol 1997;66(1):85-8. *Exclude no mass .*

Geomini P, Bremer G, Kruitwagen R, et al. Diagnostic accuracy of frozen section diagnosis of the adnexal mass: a metanalysis. [Review] [43 refs] [Journal Article. Meta-Analysis. Review] Gynecol Oncol 2005; 96(1):1-9. *Exclude review.*

Goff BA, Mandel L, Muntz HG, et al. Ovarian carcinoma diagnosis. Cancer 2000;89(10):2068-75. *Exclude review* .

Grab D, Flock F, Stohr I, et al. Classification of asymptomatic adnexal masses by ultrasound, magnetic resonance imaging, and positron emission tomography. Gynecol Oncol 2000;77(3):454-9.*Exclude Q4-not scoring system.*

Granberg S, Wikland M. A comparison between ultrasound and gynecologic examination for detection of enlarged ovaries in a group of women at risk for ovarian carcinoma. J Ultrasound Med 1988;7(2):59-64. *Exclude no mass* .

Gryspeerdt S, Clabout L, Van Hoe L, et al. Intraperitoneal contrast material combined with CT for detection of peritoneal metastases of ovarian cancer. Eur J Gynaecol Oncol 1998;19(5):434-7. *Exclude no mass* .

Guerriero S, Ajossa S, Lai MP, et al. The diagnosis of functional ovarian cysts using transvaginal ultrasound combined with clinical parameters, CA125 determinations, and color Doppler. Eur J Obstet Gynecol Reprod Biol 2003;110(1):83-8. *Exclude other all premenopausal.*

Guerriero S, Ajossa S, Lai MP, et al. Transvaginal ultrasonography associated with colour Doppler energy in the diagnosis of hydrosalpinx. Hum Reprod 2000;15(7):1568-72. *Exclude Q3-no cancer outcome.*

Guerriero S, Mais V, Ajossa S, et al. The role of endovaginal ultrasound in differentiating endometriomas from other ovarian cysts. Clin Exp Obstet Gynecol 1995;22(1):20-2. *Exclude Other All Premenopausal*.

Guerriero S, Mais V, Ajossa S, et al. Transvaginal ultrasonography combined with CA-125 plasma levels in the diagnosis of endometrioma. Fertil Steril 1996;65(2):293-8. *Exclude other all premenopausal.*

Guerriero S, Mallarini G, Ajossa S, et al. Transvaginal ultrasound and computed tomography combined with clinical parameters and CA-125 determinations in the differential diagnosis of persistent ovarian cysts in premenopausal women. Ultrasound Obstet Gynecol 1997;9(5):339-43. *Exclude Q4-all premenopausal/ Exclude Q3-all premenopausal.*

Guidozzi F. Screening for ovarian cancer. Obstet Gynecol Surv 1996;51(11):696-701. *Exclude no mass* .

Hakama M, Stenman UH, Knekt P, et al. CA 125 as a screening test for ovarian cancer. J Med Screen 1996;3(1):40-2. *Exclude no mass* .

Hall DJ, Hurt WG. The adnexal mass. J Fam Pract 1982;14(1):135-40. *Exclude no mass.*

Hamm B, Kubik-Huch RA, Fleige B. MR imaging and CT of the female pelvis: radiologic-pathologic correlation. Eur Radiol 1999;9(1):3-15. *Exclude Q3-sample size.*

Hamper UM, Sheth S, Abbas FM, et al. Transvaginal color Doppler sonography of adnexal masses: differences in blood flow impedance in benign and malignant lesions. AJR Am J Roentgenol 1993;160(6):1225-8. *Exclude Q3-unable to construct 2x2.*

Hartge P, Hayes R, Reding D, et al. Complex ovarian cysts in postmenopausal women are not associated with ovarian cancer risk factors. Am J Obstet and Gynecol 2000;183(5):1232-7. *Exclude Q1-no histol. dx.*

Hata K, Hata T, Collins WP. Association of thymidine phosphorylase concentration with ultrasound-derived indices of blood flow in ovarian masses. Cancer 1997;80(6):1079-84. *Exclude Q3-sample size*.

Hata K, Miyazaki K, Collins WP. Value of end-points from multiple or worst case Doppler spectra for the assessment of ovarian masses. Ultrasound Obstet Gynecol 1999;13(4):284. *Exclude no mass.*

Hata K, Nagami H, Iida K, et al. Expression of thymidine phosphorylase in malignant ovarian tumors: correlation with microvessel density and an ultrasound-derived index of angiogenesis. Ultrasound Obstet Gynecol 1998;12(3):201-6.
Exclude Q3-no histol. dx .

Hefler L, Mayerhofer K, Nardi A, et al. Serum soluble Fas levels in ovarian cancer. Obstet Gynecol 2000;96(1):65-9. *Exclude no mass.*

Helzlsouer KJ, Bush TL, Alberg AJ, et al. Prospective study of serum CA-125 levels as markers of ovarian cancer. JAMA 1993;269(9):1123-6. *Exclude no mass.*

Hensley ML, Castiel M, Robson ME. Screening for ovarian cancer: what we know, what we need to know. Oncology (Huntingt) 2000;14(11):1601-8, 1613-6. *Exclude no mass* .

Hricak H, Chen M, Coakley FV, et al. Complex adnexal masses: detection and characterization with MR imaging--multivariate analysis. Radiology 2000;214(1):39-46. *Exclude Q1-denom is masses.*

Hulka JF, Hulka CA. Preoperative sonographic evaluation and laparoscopic management of persistent adnexal masses: a 1994 review.... J Am Assoc Gynecol Laparosc 1994;1(3):197-205. *Exclude review.*

Im S, Gordon A, Buttin B, et al. Validation of referral guidelines for women with pelvic masses. Obstet Gyencol 2005;205(1):35-41. *Exclude Q4-not scoring system/Exclude Q3-unable to construct 2x2.*

Irwig L, Tosteson A, Gatsonis C. Guidelines for meta-analyses evaluating diagnositc tests. Ann Intern Med 1994;120:667-76. *Exclude review.*

Jacobs I. Genetic, biochemical, and multimodal approaches to screening for ovarian cancer. Gynecol Oncol 1994;55(3 Pt 2):S22-7. *Exclude no mass .*

Jacobs I, Davies AP, Bridges J, et al. Prevalence screening for ovarian cancer in postmenopausal women by CA 125 measurement and ultrasonography. BMJ 1993;306(6884):1030-4. *Exclude no mass .*

Jacobs I, Oram D. Screening for ovarian cancer. Biomed Pharmacother 1988;42(9):589-96. *Exclude review.*

Jacobs I, Stabile I, Bridges J, et al. Multimodal approach to screening for ovarian cancer. Lancet 1988;1(8580):268-71. *Exclude Q4-unable to construct 2x2.*

Jacobs IJ, Oram DH, Bast RC Jr. Strategies for improving the specificity of screening for ovarian cancer with tumor-associated antigens CA 125, CA 15-3, and TAG 72.3. Obstet Gynecol 1992;80(3 Pt 1):396-9. *Exclude no mass .*

Jacobs IJ, Rivera H, Oram DH, et al. Differential diagnosis of ovarian cancer with tumour markers CA 125, CA 15-3 and TAG 72.3. Br J Obstet Gynaecol 1993;100(12):1120-4. *Exclude Q4-partial dupl new data not relevant.*

Jain KA. Prospective evaluation of adnexal masses with endovaginal gray-scale and duplex and color Doppler US: correlation with pathologic findings. Radiology 1994;191(1):63-7. *Exclude Q1-denom is masses.*

Kadar N. Port-site recurrences following laparoscopic operations for gynaecological malignancies. Br J Obstet Gynaecol 1997;104(11):1308-13. *Exclude no mass.*

Karlan BY. Screening for ovarian cancer: what are the optimal surrogate endpoints for clinical trials? J Cell Biochem 1995;23(227-32). *Exclude review.*

Karlan BY, Platt LD. The current status of ultrasound and color Doppler imaging in screening for ovarian cancer. Gynecol Oncol 1994;55(3 Pt 2):S28-33. *Exclude review.*

Kerpsack JT, Finan MA. Thrombocytosis as a predictor of malignancy in women with a pelvic mass. J Reprod Med 2000;45(11):929-32. *Exclude Q3-wrong test.*

Kim JH, Skates SJ, Uede T, et al. Osteopontin as a potential diagnostic biomarker for ovarian cancer. JAMA 2002;287(13):1671-9. *Exclude Q3-unable to construct 2x2.*

Kinkel K, Hricak H, Lu Y, et al. US characterization of ovarian masses: a meta-analysis. Radiology 2000;217(3):803-11. *Exclude Q3-review.*

Klaren HM, van't Veer LJ, van Leeuwen FE, et al. Potential for bias in studies on efficacy of prophylactic surgery for BRCA1 and BRCA2 mutation. J Natl Cancer Inst 2003;95(13):941-7. *Exclude review.*

Kramer BS, Gohagan J, Prorok PC, et al. A National Cancer Institute sponsored screening trial for prostatic, lung, colorectal, and ovarian cancers. Cancer 1993;71(2 Suppl):589-93. *Exclude review.*

Kruitwagen RF, Swinkels BM, Keyser KG, et al. Incidence and effect on survival of abdominal wall metastases at trocar or puncture sites following laparoscopy or paracentesis in women with ovarian cancer.[see comment]. Gynecol Oncol 1996;60(2):233-7. *Exclude Q6-sample size.*

Kupesic S, Kurjak A. Contrast-enhanced, three-dimensional power Doppler sonography for differentiation of adnexal masses. Obstet Gynecol 2000;96(3):452-8. *Exclude Q3-sample size .*

Kurjak A, Kupesic S, Anic T, et al. Three-dimensional ultrasound and power doppler improve the diagnosis of ovarian lesions. Gynecol Oncol 2000;76(1):28-32. *Exclude Q3-publ duplicate.*

Kurjak A, Kupesic S, Babic MM, et al. Preoperative evaluation of cystic teratoma: what does color Doppler add? J Clin Ultrasound 1997;25(6):309-16. *Exclude no mass.*

Kurjak A, Predanic M. Ovarian cancer screening. Curr Opin Obstet Gynecol 1994;6(1):67-74. *Exclude review .*

Lang F. Resident behaviors during observed pelvic examinations. Fam Med 1990;22(2):153-5. *Exclude no mass .*

Larsen T, Torp-Pedersen ST, Ottesen M, et al. Abdominal ultrasound combined with histological and cytological fine needle biopsy of suspected ovarian tumors. Eur J Obstet Gynecol Reprod Biol 1993;50(3):203-9. *Exclude Q3-experimental or non-standard test.*

Layfield LJ, Heaps JM, Berek JS. Fine-needle aspiration cytology accuracy with palpable gynecologic neoplasms. Gynecol Oncol 1991;40(1):70-3.*Exclude Q3-sample size.*

Lee JH, Jeong YK, Park JK, et al. "Ovarian vascular pedicle" sign revealing organ of origin of a pelvic mass lesion on helical CT. AJR Am J Roentgenol 2003;181(1):131-7. *Exclude Q3-distguish malignant versus nonmalignant.*

Lehner R, Wenzl R, Heinzl H, et al. Influence of delayed staging laparotomy after laparoscopic removal of ovarian masses later found malignant. Obstet Gynecol 1998;92(6):967-71. *Exclude no mass .*

Lerner JP, Timor-Tritsch IE, Federman A, et al. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. Am J Obstet Gynecol 1994;170(1 Pt 1):81-5. *Exclude Q1-denom is masses.*

Levine D, Feldstein VA, Babcook CJ, et al. Sonography of ovarian masses: poor sensitivity of resistive index for identifying malignant lesions. AJR Am J Roentgenol 1994;162(6):1355-9. *Exclude Q3-sample size.*

Levine D, Gosink BB, Wolf SI, et al. Simple adnexal cysts: the natural history in postmenopausal women.[see comment]. Radiology 1992;184(3):653-9. *Exclude Q6-sample size.*

Levy G, Levine P, Brennan J, et al. Color flow-directed Doppler studies of ovarian masses. Computer analysis. J Reprod Med 1998;43(10):865-8. *Exclude Other.*

Lieberman G, Buscombe JR, Hilson AJ, et al. Preoperative diagnosis of ovarian carcinoma with a novel monoclonal antibody. Am J Obstet Gynecol 2000;183(3):534-40. *Exclude Q3-experimental or non-standard test.*

Lieberman G, MacLean AB, Buscombe JR, et al. The clinical application of a dual head gamma camera with coincidence detection in 20 women with suspected ovarian cancer. BJOG 2001;108(12):1229-36. *Exclude no mass*.

Liede A, Karlan B, Baldwin RL, et al. Cancer Incidence in a population of jewish women at risk of ovarian cancer. J Clin Oncol 2002;20(6):1570-77. *Exclude review*.

Lin P, Falcone T, Tulandi T. Excision of ovarian dermoid cyst by laparoscopy and by laparotomy. Am J Obstet Gynecol 1995;173(3 Pt 1):769-71. *Exclude Q6-sample size.*

Lynch HT, Albano WA, Lynch JF, et al. Surveillance and management of patients at high genetic risk for ovarian carcinoma. Obstet Gynecol 1982;59(5):589-96. *Exclude no mass* .

MacDonald ND, Rosenthal AN, Jacobs IJ. Screening for ovarian cancer. Ann Acad Med Singapore 1998;27(5):676-82. *Exclude review.*

Mackey SE, Creasman WT. Ovarian cancer screening. J Clin Oncol 1995;13(3):783-93. *Exclude review.*

Maggino T, Gadducci A. Serum markers as prognostic factors in epithelial ovarian cancer: an overview. Eur J Gynaecol Oncol 2000;21(1):64-9. *Exclude review.*

Mais V, Ajossa S, Piras B, et al. Treatment of nonendometriotic benign adnexal cysts: a randomized comparison of laparoscopy and laparotomy. Obstet Gynecol 1995;86(5):770-4. *Exclude Q6-sample size.*

Mais V, Guerriero S, Ajossa S, et al. Transvaginal ultrasonography in the diagnosis of cystic teratoma. Obstet Gynecol 1995;85(1):48-52. *Exclude Q3-all premenopausal.*

Markman M. Limitations to the use of the CA-125 antigen level in ovarian cancer. Curr Oncol Rep 2003;5(4):263-4. *Exclude review.*

Masson V. Bodies of knowledge. Nurs Health Care Perspect 1997;18(6):291. *Exclude no mass* .

McIntosh MW, Urban N. A parametric empirical Bayes method for cancer screening using longitudinal observations of a biomarker. Biostatistics 2003;4(1):27-40. *Exclude Q7-not descrip of sim model.*

Mendilcioglu I, Zorlu CG, Trak B, et al. Laparoscopic management of adnexal masses. Safety and effectiveness. J Reprod Med 2002;47(1):36-40. *Exclude Q6-sample size.*

Menon U, Jacobs IJ. Ovarian cancer screening in the general population. Curr Opin Obstet Gynecol 2001;13(1):61-4. *Exclude review.*

Menon U, Talaat A, Rosenthal AN, et al. Performance of ultrasound as a second line test to serum CA125 in ovarian cancer screening. BJOG 2000;107(2):165-9. *Exclude Q4-wrong pt population.*

Mensah LG. Gynecology in the generalist's office. Ethn Dis 2003;13(3 Suppl 3):S3-50-1. *Exclude Q1-no histol. dx.*

Merz E, Miric-Tesanic D, Bahlmann F, et al. Sonographic size of uterus and ovaries in pre- and postmenopausal women. Ultrasound Obstet Gynecol 1996;7(1):38-42. *Exclude no mass* .

Mettler L, Semm K, Shive K. Endoscopic management of adnexal masses. J Soc Laparoendosc Surg 1997;1(2):103-12. *Exclude Q6-no M&M data.*

Midgette AS, Stukel TA , Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. Med Decis Making 1993;13:253-7. *Exclude review.*

Milad MP, Cohen L. Preoperative ultrasound assessment of adnexal masses in premenopausal women. Int J Gynaecol Obstet 1999;66(2):137-41. *Exclude other all premenopausal.*

Mills GB, Bast RC Jr, Srivastava S. Future for ovarian cancer screening: novel markers from emerging technologies of transcriptional profiling and proteomics. J Natl Cancer Inst 2001;93(19):1437-9. *Exclude review*.

Misawa T, Asai M, Higashide K. How to decrease false-positive cases of ovarian cancer screening by transvaginal sonography. J Exp Clin Cancer Res 1997;16(2):217-20. *Exclude no mass* .

Modesitt SC, Pavlik EJ, Ueland FR, et al. Risk of malignancy in unilocular ovarian cystic tumors less than 10 centimeters in diameter. Obstet Gynecol 2003;102(3):594-9. *Exclude Q3-no cancer outcome*.

Morgan A. Adnexal mass evaluation in the emergency department. Emerg Med Clin North Am 2001;19(3):799-816. *Exclude review.*

Moses LE, Shapiro D. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993;12:1293-316.*Exclude review.*

Muto MG, Cramer DW, Brown DL, et al. Screening for ovarian cancer: the preliminary experience of a familial ovarian cancer center. Gynecol Oncol 1993;51(1):12-20. *Exclude Q3-sample size.*

Nagarsheth NP, Rahaman J, Cohen CJ, et al. The incidence of port-site metastases in gynecologic cancers. J Soc Laparoendosc Surg 2004;8(2):133-9.*Exclude no mass.*

Nevin J, Denny L, Soeters R, et al. Ultrasonography of pelvic masses. Br J Obstet Gynaecol 1998;105(2):137-9. *Exclude review.*

Nezhat C, Santolaya J, Nezhat FR. Comparison of transvaginal sonography and bimanual pelvic examination in patients with laparoscopically confirmed endometriosis. J Am Assoc Gynecol Laparosc 1994;1(2):127-30. *Exclude Q3-distguish malignant versus nonmalignant.*

Nezhat F, Nezhat C, Welander CE, et al. Four ovarian cancers diagnosed during laparoscopic management of 1011 women with adnexal masses. Am J Obstet Gynecol 1992;167(3):790-6. *Exclude other.*

Nichols M, Morgan E, Jensen JT. Comparing bimanual pelvic examination to ultrasound measurement for assessment of gestational age in the first trimester of pregnancy. J Reprod Med 2002;47(10):825-8. *Exclude Q1-no histol. dx.*

O'Rourke J, Mahon SM. A comprehensive look at the early detection of ovarian cancer. Clin J Oncol Nurs 2003;7(1):41-7. *Exclude no mass* .

Ong S, Duffy T, Murphy J. Transabdominal ultrasound and its correlation with clinical findings in gynaecology. Ir J Med Sci 1996;165(4):268-70. *Exclude Q3-unable to construct 2x2.*

Onsrud M, Shabana A, Austgulen R, et al. Comparison between soluble tumor necrosis factor receptors and CA125 in peritoneal fluids as a marker for epithelial ovarian cancer. Gynecol Oncol 1995;57(2):183-7. *Exclude Q3-wrong test.*

Opala T, Drews K, Rzymski P, et al. Evaluation of soluble intracellular adhesion molecule-1 (sICAM-1) in benign and malignant ovarian masses. Eur J Gynaecol Oncol 2003;24(3-4):255-7. *Exclude Q3-experimental or non-standard test.*

Opala T, Rzymski P, Wilczak M, et al. Evaluation of soluble tumour necrosis factor alpha receptors p55 and p75 in ovarian cancer patients. Eur J Gynaecol Oncol 2005;26(1):43-6.*Exclude Q3-wrong test.*

Oram DH, Jacobs IJ, Brady JL, et al. Early diagnosis of ovarian cancer. Br J Hosp Med 1990;44(5):320, 322, 324. *Exclude Q3-review*.

Paley PJ. Screening for the major malignancies affecting women: current guidelines. Am J Obstet Gynecol 2001;184(5):1021-30.*Exclude review.*

Papasakelariou C, Saunders D, De La Rosa A. Comparative study of laparoscopic oophorectomy. J Am Assoc Gynecol Laparosc 1995;2(4):407-10.*Exclude Q6-no M&M data.*

Pardo J, Kaplan B, Yitzhak M, et al. Ultrasonographic evaluation of hysterectomized patients with and without concomitant adnexectomy? Clin Exp Obstet Gynecol 1998;25(4):133-4. *Exclude no mass* .

Parker WH. Management of adnexal masses by operative laparoscopy. Selection criteria. J Reprod Med 1992;37(7):603-6. Exclud*e no mass.*

Parker WH, Berek JS. Management of selected cystic adnexal masses in postmenopausal women by operative laparoscopy: a pilot study. Am J Obstet Gynecol 1990;163(5 Pt 1):1574-7. *Exclude Q2-no physical exam.*

Patel MD, Feldstein VA, Chen DC, et al. Endometriomas: diagnostic performance of US. [erratum appears in Radiology 1999 Dec;213(3):930]. Radiology 1999;210(3):739-45. *Exclude Q3-distguish malignant versus nonmalignant.*

Patel MD, Feldstein VA, Lipson SD, et al. Cystic teratomas of the ovary: diagnostic value of sonography. AJR Am J Roentgenol 1998;171(4):1061-5. *Exclude Q3-distguish malignant versus nonmalignant*.

Patsner B, Mann WJ, Chalas E. Predictive value of CA 125 for ovarian carcinoma in patients presenting with pelvic masses. Obstet Gynecol 1988;71(6 Pt 1):949-50. *Exclude review.*

Pauler DK, Menon U, McIntosh M, et al. Factors influencing serum CA125II levels in healthy postmenopausal women. Cancer Epidemiol Biomarkers Prev 2001;10(5):489-93. *Exclude Q1-no histol. dx.*

Pearson VA. Screening for ovarian cancer: a review.... Public Health 1994;108(5):367-82. *Exclude review .*

Petricoin EF, Ardekani AM, Hitt BA, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359(9306):572-7.*Exclude no mass.*

Predanic M, Vlahos N, Pennisi JA, et al. Color and pulsed Doppler sonography, gray-scale imaging, and serum CA 125 in the assessment of adnexal disease. Obstet Gynecol 1996;88(2):283-8. *Exclude Q1-sample size.*

Prefontaine M, Kroft T, Monck M, et al. Evaluation of a simple line width test involving magnetic resonance spectroscopy of plasma in carcinoma of the ovary. Cancer 1991;67(2):406-11.*Exclude Q3-wrong test .*

Pugh CM, Youngblood P. Development and validation of assessment measures for a newly developed physical examination simulator. J Am Med Inform Assoc 2002;9(5):448-60. *Exclude review.*

Qureshi IA, Ullah H, Akram MH, et al. Transvaginal versus transabdominal sonography in the evaluation of pelvic pathology. J Coll Physicians Surg Pak 2004;14(7):390-3. *Exclude Q3-no histol. dx .*

Ramirez PT, Frumovitz M, Wolf JK, et al. Laparoscopic port-site metastases in patients with gynecological malignancies. [Review...] [58 refs]. Int J Gynecol Cancer 2004;14(6):1070-7.*Exclude review.*

Rifkin JI, Shapiro H, Regensteiner JG, et al. Why do some women refuse to allow male residents to perform pelvic exams? Acad Med 2002;77(10):1034-8. *Exclude no mass .*

Rollins G. Developments in cervical and ovarian cancer screening: implications for current practice. Ann Intern Med 2000;133(12):1021-4. *Exclude review.*

Roman LD, Muderspach LI, Stein SM, et al. Pelvic examination, tumor marker level, and gray-scale and Doppler sonography in the prediction of pelvic cancer. Obstet Gynecol 1997;89(4):493-500. *Exclude Q7-not descrip of sim model/Exclude Q5-wrong pt population.*

Rosenthal AN, Jacobs IJ. The role of CA 125 in screening for ovarian cancer. Int J Biol Markers 1998;13(4):216-20. *Exclude review.*

Sari R, Buyukberber S, Sevinc A, et al. The effects of abdominal and bimanual pelvic examination and transvaginal ultrasonography on serum CA-125 levels. Clin Exp Obstet Gynecol 2000;27(1):69-71. *Exclude review.*

Sassone AM, Timor-Tritsch IE, Artner A, et al. Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy. Obstet Gynecol 1991;78(1):70-6. *Exclude Q1-denom is masses.*

Schwartz LB, Seifer DB. Diagnostic imaging of adnexal masses. A review. J Reprod Med 1992;37(1):63-71. *Exclude review .*

Seewaldt VL, Cain JM, Greer BE, et al. Reviving the pelvic examination for evaluating the status of ovarian carcinoma. J Clin Oncol 1995;13(3):799. *Exclude Q1-no histol. dx.*

Sengoku K, Satoh T, Saitoh S, et al. Evaluation of transvaginal color Doppler sonography, transvaginal sonography and CA 125 for prediction of ovarian malignancy. Int J Gynaecol Obstet 1994;46(1):39-43. *Exclude Q4-sample size.*

Shaharabany Y, Akselrod S, Tepper R. A sensitive new indicator for diagnostics of ovarian malignancy, based on the Doppler velocity spectrum. Ultrasound Med Biol 2004;30(3):295-302 *Exclude Q3-unable to construct 2x2.*

Shalev E, Eliyahu S, Peleg D, et al. Laparoscopic management of adnexal cystic masses in postmenopausal women. Obstet Gynecol 1994;83(4):594-6. *Exclude Q3-not complete series.*

Shapiro I, Friedman Z, Lysyansky P, et al. The instantaneous measurement of multiple Doppler spectra in the investigation of ovarian masses. Ultrasound Obstet Gynecol 1998;11(5):353-6. *Exclude Q3-no histol. dx*

Shen-Gunther J, Mannel RS. Ascites as a predictor of ovarian malignancy. Gynecol Oncol 2002;87(1):77-83. *Exclude Q3-unable to construct 2x2/ Exclude Q2-only 1 patient had ascites.*

Sheppard R, Fry A, Rush R, et al. Women at risk of ovarian cancer: attitudes towards and expectations of the familial ovarian cancer clinic. Fam Cancer 2001;1(1):31-7. *Exclude no mass.*

SOGC/GOC/SCC Policy and Practice Guideline Committee. Guidelines for the Laparoscopic Management of the adnexal mass: a policy statement. SOGC Clinical Practice Guidelines 1998;(76). *Exclude review.*

Soriano D, Yefet Y, Seidman DS, et al. Laparoscopy versus laparotomy in the management of adnexal masses during pregnancy. Fertil Steril 1999;71(5):955-60. *Exclude Other All Premenopausal.*

Sparks JM, Varner RE. Ovarian cancer screening. Obstet Gynecol 1991;77(5):787-92. *Exclude review.*

Stein SM, Laifer-Narin S, Johnson MB, et al. Differentiation of benign and malignant adnexal masses: relative value of gray-scale, color Doppler, and spectral Doppler sonography. AJR Am J Roentgenol 1995;164(2):381-6. *Exclude Q1-denom is masses.*

Tailor A, Jurkovic D, Bourne TH, et al. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. Ultrasound Obstet Gynecol 1997;10(1):41-7. *Exclude Q4-sample size.*

Tailor A, Jurkovic D, Bourne TH, et al. Sonographic prediction of malignancy in adnexal masses using an artificial neural network. Br J Obstet Gynaecol 1999;106(1):21-30. *Exclude Q3-unable to construct 2x2.*

Tailor A, Jurkovic D, Bourne TH, et al. A comparison of intratumoural indices of blood flow velocity and impedance for the diagnosis of ovarian cancer. Ultrasound Med Biol 1996;22(7):837-43. *Exclude Q3-sample size*.

Tangjitgamol S, Jesadapatrakul S, Manusirivithaya S, et al. Accuracy of frozen section in diagnosis of ovarian mass. Int J Gynecol Cancer 2004;14(2):212-9. *Exclude Q3-wrong test.*

Tekay A, Jouppila P. Controversies in assessment of ovarian tumors with transvaginal color Doppler ultrasound. Acta Obstet Gynecol Scand 1996;75(4):316-29. *Exclude Q3-review.*

Tempfer C, Hefler L, Heinzl H, et al. CYFRA 21-1 serum levels in women with adnexal masses and inflammatory diseases. Br J Cancer 1998;78(8):1108-12. *Exclude Q3-experimental or non-standard test.*

Teneriello MG, Park RC. Early detection of ovarian cancer. CA Cancer J Clin 1995;45(2):71-87. *Exclude review.*

Tepper R, Keselbrener L, Manor M, et al. Decay constant of Doppler flow waveform as a possible indicator of ovarian malignancy. Ultrasound Med Biol 1997;23(8):1171-7. *Exclude Q3-unable to construct 2x2.*

Thorvinger B. Diagnostic and interventional radiology in gynecologic neoplasms. Acta Radiol Suppl 1992;378(Pt 3):93-108. *Exclude Q1-no histol. dx.*

Timmerman D, Bourne TH, Tailor A, et al. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model. Am J Obstet Gynecol 1999;181(1):57-65. *Exclude Q3-unable to construct 2x2.*

Timmerman D, Schwarzler P, Collins WP, et al. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. Ultrasound Obstet Gynecol 1999;13(1):11-6. *Exclude Q3-unable to construct 2x2.*

Timor-Tritsch IE. Is office use of vaginal ultrasonography feasible? Am J Obstet Gynecol 1990;162(4):983-5. *Exclude Q1-no histol. dx.*

Togashi K, Nishimura K, Kimura I, et al. Endometrial cysts: diagnosis with MR imaging. Radiology 1991;180(1):73-8.
Exclude Q3-no cancer outcome.

Tuxen MK, Soletormos G, Dombernowsky P. Tumor markers in the management of patients with ovarian cancer. Cancer Treat Rev 1995;21(3):215-45. *Exclude Q3-review.*

U.S. Preventive Services Task Force. Screening for ovarian cancer: recommendation statement. Ann Fam Med 2004;2(3):260-2. *Exclude Q1-no histol. dx.*

Ueland F, DePriest P, DeSimone C, et al. The accuracy and examination under anesthesia and transvaginal sonography in evaluating ovarian size. Gynecol Oncol 2005 (in Press) *Exclude no mass.*

Urban N. Screening for ovarian cancer. We now need a definitive randomised trial. BMJ 1999;319(7221):1317-8.
Exclude Q1-no histol. dx.

Urban N, McIntosh MW, Andersen M, et al. Ovarian cancer screening. Hematol Oncol Clin North Am 2003;17(4):985-1005. *Exclude review.*

Usubutun A, Altinok G, Kucukali T. The value of intraoperative consultation (frozen section) in the diagnosis of ovarian neoplasms. Acta Obstet Gynecol Scand 1998;77(10):1013-6. *Exclude Q3-wrong test.*

Vaidya AP, Curtin JP. The follow-up of ovarian cancer. Semin Oncol 2003;30(3):401-12. *Exclude Q1-no histol. dx.*

Valentin L. Prospective cross-validation of Doppler ultrasound examination and gray-scale ultrasound imaging for discrimination of benign and malignant pelvic masses. Ultrasound Obstet Gynecol 1999;14(4):273-83.
Exclude Q3-publ duplicate.

Valentin L, Sladkevicius P, Marsal K. Limited contribution of Doppler velocimetry to the differential diagnosis of extrauterine pelvic tumors. Obstet Gynecol 1994;83(3):425-33. *Exclude Q3-unable to construct 2x2.*

van Dam PA, DeCloedt J, Tjalma WA, et al. Trocar implantation metastasis after laparoscopy in patients with advanced ovarian cancer: can the risk be reduced?. Am J Obstet Gynecol 1999;181(3):536-41. *Exclude no mass*

van der Burg ME, Lammes FB, Verweij J. The role of CA 125 and conventional examinations in diagnosing progressive carcinoma of the ovary. Surg Gynecol Obstet 1993;176(4):310-4. *Exclude Q3-wrong test.*

van Nagell JR Jr, DePriest PD, Gallion HH, et al. Ovarian cancer screening. Cancer 1993;71(4 Suppl):1523-8. *Exclude review* .

van Nagell JR Jr, DePriest PD, Reedy MB, et al. The efficacy of transvaginal sonographic screening in asymptomatic women at risk for ovarian cancer. Gynecol Oncol 2000;77(3):350-6. *Exclude Q5-wrong pt population.*

van Nagell JR Jr, Higgins RV, Donaldson ES, et al. Transvaginal sonography as a screening method for ovarian cancer. A report of the first 1000 cases screened. Cancer 1990;65(3):573-7.*Exclude Q1-pop is subset of larger study (#2730).*

van Nagell JR Jr, Ueland FR. Ultrasound evaluation of pelvic masses: predictors of malignancy for the general gynecologist. Curr Opin Obstet Gynecol 1999;11(1):45-9.*Exclude review*.

Varpula M. Magnetic resonance imaging of female pelvic masses and local recurrent tumors at an ultra low (0.02 T) magnetic field: correlation with computed tomography. Magn Reson Imaging 1993;11(1):35-46.*Exclude Q3-sample size*

Varras M. Benefits and limitations of ultrasonographic evaluation of uterine adnexal lesions in early detection of ovarian cancer. Clin Exp Obstet Gynecol 2004;31(2):85-98. *Exclude review.*

Vogl FD, Frey M, Kreienberg R, et al. Autoimmunity against p53 predicts invasive cancer with poor survival in patients with an ovarian mass. Br J Cancer 2000;83(10):1338-43. *Exclude Q3-sample size.*

von Schlippe M, Rustin GJ. Circulating tumour markers in ovarian tumours. Forum 2000;10(4):383-92. *Exclude review.*

Voss SC, Lacey CG, Pupkin M, et al. Ultrasound and the pelvic mass. J Reprod Med 1983;28(12):833-7. Exclude Q3-no verification test negative/ *Exclude Q2-unable to construct 2x2.*

Wakahara F, Kikkawa F, Nawa A, et al. Diagnostic efficacy of tumor markers, sonography, and intraoperative frozen section for ovarian tumors. Gynecol Obstet Invest 2001;52(3):147-52. *Exclude Q4-unable to construct 2x2.*

Wardle FJ, Collins W, Pernet AL, et al. Psychological impact of screening for familial ovarian cancer. J Natl Cancer Inst 1993;85(8):653-7. *Exclude review.*

Weerakiet S, Wongkularb A, Rochanawutanon M, et al. Transvaginal ultrasonography combined with pelvic examination in the diagnosis of ovarian endometrioma. J Med Assoc Thai 2000;83(5):523-8. *Exclude Q3-no cancer outcome.*

Weiner Z, Beck D, Brandes JM. Transvaginal sonography, color flow imaging, computed tomographic scanning, and CA 125 as a routine follow-up examination in women with pelvic tumor: detection of recurrent disease. J Ultrasound Med 1994;13(1):37-41. *Exclude no mass* .

Westhoff C, Levin B, Ladd G, et al. Sources of variability in normal CA 125 levels. Cancer Epidemiol Biomarkers Prev 1992;1(5):357-9. *Exclude review.*

Yawn BP, Wollan PC. Ovarian cancer: the neglected diagnosis. Mayo Clinic Proceedings 2004;79(10):1277-82. *Exclude review.*

Zhang Z, Bast RC Jr, Yu Y, et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer.[see comment]. Cancer Res 2004;64(16):5882-90. *Exclude Q3-experimental or non-standard test.*

Zurawski VR Jr, Knapp RC, Einhorn N, et al. An initial analysis of preoperative serum CA 125 levels in patients with early stage ovarian carcinoma. Gynecol Oncol 1988;30(1):7-14. *Exclude Q3-distguish malignant versus nonmalignant.*

Zygmunt A, Markowska J, Fischer N. Significance of tissue polypeptide specific antigen (TPS) in diagnosis and monitoring of treatment in ovarian cancer. Eur J Gynaecol Oncol 1998;19(5):484-6. *Exclude review.*

Appendix C: Sample Data Abstraction Forms

Question 1: *What is the prevalence of various tumor types among peri- and postmenopausal women with an adnexal mass, stratified by cancer status (malignant vs. benign), age, and size of tumor?*

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
StudyID	**Geographical location:**	**Age:** Mean (SD): Median: Range:	**Symptomatic (n [%]):**	[Proportion of each type of finding, stratified by cancer status, age/menopausal status (<45, 45-55, >55 or pre-peri-post-menopausal), and size of tumor. Include individual tumor types where possible.]	[IF ARTICLE SHOULD BE EXCLUDED, PLEASE EXPLAIN WHY HERE]
	Dates:		**Detected by exam (n [%]):**		
		Menopausal status (n [%]):		Use Excel spreadsheet to calculate confidence intervals for prevalence data from screening studies	
	Size of population: [num/denom for screening studies]	Pre (< 45): Peri (45-55): Post (> 55):	**Detected by imaging (n [%]):**	1)	[COMMENT ON BIASES, ETC. AFFECTING CLINICAL INTERPRETATION]
		Race/ethnicity (n [%]):	**Combination (n [%]):**		
	Screening study **Registry** **Other** [delete all but one; please specify "Other"]			2)	**Quality assessment:** [assign + or - to each item, and provide a brief rationale]
		Risk factors (n [%]): Family history: Genotype: Other [specify]:	**Additional data used for diagnosis:**	3)	Size of population from which sample drawn: Number of cases: Patient selection: Application of reference standard:
				4)	
					This article is also relevant to: [delete as appropriate]
				5)	Question 2 Question 3 Question 4 Question 5 Question 6 Question 7

Question 2: *What are the sensitivity, specificity, and reliability of the bimanual examination?*

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
StudyID	**Geographical location:** **Dates:** **Size of population:** [num/denom for screening studies] **Screening study** **Registry** **Other** [delete all but one; please specify "Other"] **Reference standard:** **Reference standard applied to all test negatives?:** **Test reliability established?:** **Statistical tests used:** **Blinding:** **Definition of positive and negative on screening test:**	**Age:** Mean (SD): Median: Range: **Menopausal status (n [%]):** Pre (< 45): Peri (45-55): Post (> 55): **Race/ethnicity (n [%]):** **Risk factors (n [%]):** Family history: Genotype: Other [specify]: **Inclusion criteria:** **Exclusion criteria:**	**Symptomatic (n [%]):** **Detected by exam (n [%]):** **Detected by imaging (n [%]):** **Combination (n [%]):** **Additional data used for diagnosis:**	[Please provide brief description of clinical setting in which bimanual exam was performed]	[For bimanual exam, provide reported sensitivity/specificity and provide 2x2 tables (if possible). If possible and appropriate, stratify by age or menopausal status. If data are available on reliability/ reproducibility, report these as well. Include kappa scores if these are reported or can be calculated.] 1) [Use this space to provide information needed for reader to interpret Test +, Test -, Disease +, and Disease - headings in following table.] 2)	[IF ARTICLE SHOULD BE EXCLUDED, PLEASE EXPLAIN WHY HERE] [COMMENT ON BIASES, ETC. AFFECTING CLINICAL INTERPRETATION] **Quality assessment:** [assign + or - to each item, and provide a brief rationale] Reference standard: Verification bias: Test reliability/variability: Sample size: Statistical tests: Blinding: Definition of +/- on screening test: **This article is also relevant to:** [delete as appropriate] Question 1 Question 2 Question 3 Question 4 Question 5 Question 7

Question 3: *Among peri- and postmenopausal women with a palpable adnexal mass on exam or a mass identified by ultrasound/imaging, what is the sensitivity/specificity of various evaluation modalities including ultrasound (transvaginal ultrasound, transabdominal ultrasound, color Doppler, 2-D vs 3D ultrasound, CT scan, MRI scan, and CA-125 levels) for diagnosing malignant masses?*

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
StudyID	**Geographical location:** **Dates:** **Size of population:** [num/denom for screening studies] **Screening study** **Registry** **Other** [delete all but one; please specify "Other"] **Reference standard:** **Reference standard applied to all test negatives?:** **Test reliability established?:** **Statistical tests used:** **Blinding:** **Definition of positive and negative on screening test:**	**Age:** Mean (SD): Median: Range: **Menopausal status (n [%]):** Pre (< 45): Peri (45-55): Post (> 55): **Race/ethnicity (n [%]):** **Risk factors (n [%]):** Family history: Genotype: Other [specify]: **Inclusion criteria:** **Exclusion criteria:**	**Symptomatic (n [%]):** **Detected by exam (n [%]):** **Detected by imaging (n [%]):** **Combination (n [%]):** **Additional data used for diagnosis:**	[For each test reported, please provide a 2x2 table and report or calculate sensitivity, specificity, NPV, and PPV (all with confidence intervals). If possible and appropriate, stratify by age or menopausal status.] 1) [Use this space to provide information needed for reader to interpret Test +, Test -, Disease +, and Disease - headings in following table.] 2)	[IF ARTICLE SHOULD BE EXCLUDED, PLEASE EXPLAIN WHY HERE] [COMMENT ON BIASES, ETC. AFFECTING CLINICAL INTERPRETATION] **Quality assessment:** [assign + or - to each item, and provide a brief rationale] Reference standard: Verification bias: Test reliability/variability: Sample size: Statistical tests: Blinding: Definition of +/- on screening test: **This article is also relevant to:** [delete as appropriate] Question 1 Question 3 Question 4 Question 5 Question 6 Question 7

Question 4: *What is the accuracy of explicit scoring systems which incorporate various combinations of imaging findings, patient risk factors, and/or CA-125 levels for detecting malignancy? Have these scoring systems been applied to a population of peri-/postmenopausal women before laparoscopy*?

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
StudyID	**Geographical location:** **Dates:** **Size of population:** [num/denom for screening studies] **Screening study** **Registry** **Other** [delete all but one; please specify "Other"] **Reference standard:** **Reference standard applied to all test negatives?:** **Statistical tests used:** **Blinding:** **Definition of positive and negative on screening test:**	**Age:** Mean (SD): Median: Range: **Menopausal status (n [%]):** Pre (< 45): Peri (45-55): Post (> 55): **Race/ethnicity (n [%]):** **Risk factors (n [%]):** Family history: Genotype: Other [specify]: **Inclusion criteria:** **Exclusion criteria:**	**Symptomatic (n [%]):** **Detected by exam (n [%]):** **Detected by imaging (n [%]):** **Combination (n [%]):** **Additional data used for diagnosis:**	1) 2) 3) 4) 5) 6) 7) 8) 9) 10)	[For each reported scoring system (and individual components, if reported), provide reported sensitivity/specificity and provide 2x2 table; if multivariate analysis, provide area under ROC curve or c-statistic, if reported. If possible and appropriate, stratify by age or menopausal status.] 1) [Use this space to provide information needed for reader to interpret Test +, Test -, Disease +, and Disease - headings in following table.] 2)	[IF ARTICLE SHOULD BE EXCLUDED, PLEASE EXPLAIN WHY HERE] [COMMENT ON BIASES, ETC. AFFECTING CLINICAL INTERPRETATION] **Quality assessment:** [assign + or - to each item, and provide a brief rationale] Reference standard: Verification bias: Test reliability/variability: Sample size: Statistical tests: Blinding: Definition of +/- on screening test: Explicit validation method?: **This article is also relevant to:** [delete as appropriate] Question 1 Question 2 Question 4 Question 5 Question 6 Question 7

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
					Results were reported, but have not been abstracted, for the following combinations: [list]	

Question 5: *Among women with suspected benign lesions on initial investigation, what is the sensitivity and specificity of monitoring with periodic CA-125 and/or interval ultrasound examinations for detecting malignant masses? How does the interval of testing/definition of change affect sensitivity and predictive value?*

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
StudyID	**Geographical location:** **Dates:** **Size of population:** [num/denom for screening studies] **Screening study** **Registry** **Other** [delete all but one; please specify "Other"] **Reference standard:** **Reference standard applied to all test negatives?:** **Test reliability established?:** **Statistical tests used:**	**Age:** Mean (SD): Median: Range: **Menopausal status (n [%]):** Pre (< 45): Peri (45-55): Post (> 55): **Race/ethnicity (n [%]):** **Risk factors (n [%]):** Family history: Genotype: Other [specify]: **Inclusion criteria:** **Exclusion criteria:** **Loss to follow up:**	**Symptomatic (n [%]):** **Detected by exam (n [%]):** **Detected by imaging (n [%]):** **Combination (n [%]):** **Additional data used for diagnosis:**	**Monitoring test:** **Interval of testing:** **Definition of change:**	[For each reported monitoring strategy, provide reported sensitivity/specificity and provide 2x2 table; if multivariate analysis, provide area under ROC curve or c-statistic, if reported. If possible and appropriate, stratify by age or menopausal status.] 1) [Use this space to provide information needed for reader to interpret Test +, Test -, Disease +, and Disease - headings in following table.] 2)	[IF ARTICLE SHOULD BE EXCLUDED, PLEASE EXPLAIN WHY HERE] [COMMENT ON BIASES, ETC. AFFECTING CLINICAL INTERPRETATION] **Quality assessment:** [assign + or - to each item, and provide a brief rationale] Reference standard: Verification bias: Test reliability/variability: Sample size: Statistical tests: Blinding: Definition of +/- on screening test: Explicit validation method?: **This article is also relevant to:** [delete as appropriate]

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
	Blinding:					Question 1 Question 2 Question 3 Question 5 Question 6 Question 7
	Definition of positive and negative on screening test:					
	Length of follow up:				3)	
	Type of follow up:					
	Follow-up interval:					

Question 6: *Among women with adnexal masses, what is the morbidity and mortality from diagnostic laparoscopy? At what point does the risk of laparoscopy outweigh the risk of detecting malignancy?*

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
StudyID	**Geographical location:**	**Age:** Mean (SD): Median: Range:	**Symptomatic (n [%]):**	[For each, provide reported rate and 95% CI, if appropriate. If possible and appropriate, stratify results by age or menopausal status.]	[IF ARTICLE SHOULD BE EXCLUDED, PLEASE EXPLAIN WHY HERE]
	Dates:		**Detected by exam (n [%]):**	Use Excel spreadsheet to calculate confidence intervals for morbidity/mortality	
		Menopausal status (n [%]):			[COMMENT ON BIASES, ETC. AFFECTING CLINICAL INTERPRETATION]
	Size of population: [num/denom for screening studies]	Pre (< 45): Peri (45-55): Post (> 55):	**Detected by imaging (n [%]):**	1) Mortality:	
		Race/ethnicity (n [%]):	**Combination (n [%]):**		
	Single center **Registry** [delete one]			2) Morbidity (total all complications):	**Quality assessment:** [assign + or - to each item, and provide a brief rationale]
		Risk factors (n [%]): Family history: Genotype:	**Additional data used for diagnosis:**	3) Specific complications:	Size of population from which sample drawn: Number of cases: Patient selection: Application of reference standard:
	Morbidity definitions:	Other [specify]:			
	Length of follow up after surgery:	**Loss to follow up:**		4) Rate of conversion to laparotomy:	
				5)	**This article is also relevant to:** [delete as appropriate] Question 1 Question 2 Question 3 Question 4
				6)	Question 6 Question 7

Question 7: *What are the estimated trade-offs resulting from various strategies for evaluation of the adnexal mass?*

Study	Study Design	Study Outcomes	Sources for Model Probabilities	Sources for Model Outcomes	Results	Comments
StudyID	**Type of model:** **Population modeled (age, range):** **Strategies compared**:	[Life expectancy, quality of life, cancer incidence, cancer death, etc. Include costs, but we will not be using them here]	[In particular, sources for transition probabilities between different stages of pre-cancer/cancer] **Simplifying assumptions:**		[For each strategy compared, compare results for different outcomes; also, report results of significant sensitivity analyses.] 1) 2) 3) 4) 5) 6)	[IF ARTICLE SHOULD BE EXCLUDED, PLEASE EXPLAIN WHY HERE] [COMMENT ON BIASES, ETC. AFFECTING CLINICAL INTERPRETATION] **This article is also relevant to:** [delete as appropriate] Question 1 Question 2 Question 3 Question 4 Question 5 Question 6

Appendix D: Evidence Tables

Evidence Table 1: *Question 1: What is the prevalence of various tumor types among women with an adnexal mass, stratified by cancer status (malignant vs. benign), age, menopausal status, and size of tumor?*

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Chalas, Welshinger, Engellener, et al., 1992 **#5100**	**Geographical location:** Stony Brook, NY **Dates:** May 1980-Apr 1990 **Size of population:** 241 **Other:** Retrospective chart review of patients with pelvic mass who underwent laparotomy to look at thrombocytosis as a predictor of cancer	**Age:** NR **Menopausal status (n [%]):** NR; authors present some findings by age > 50, but do not report the numbers of women **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** CA-125 and thrombocytosis	Of the 241: 121/241 were malignant = 50.2%; 95% CI, 44.4 to 56.9 18/241 borderline = 7.5%; 4.8 to 11.7 102/241 benign = 42.3%; 36.6 to 49 **Malignant:** Serous epithelial 66 = 27.4%; 95% CI, 22.4 to 33.6 Mucinous epithelial 12 = 5.0%; 2.8 to 8.7 Clear cell epithelial 13 = 5.4%; 3.2 to 9.2 Endometrioid epithelial 11 = 4.6%; 2.5 to 8.2 Papillary mixed epithelial 5 = 2.1%; 0.8 to 5.0 Dysgerminoma 2 = 0.8%; 0.05 to 3.2 Immature teratoma 1 = 0.4%; 0 to 2.6 Endodermal tumor 1 = 0.4%; 0 to 2.6 Granulose cell tumor 1 = 0.4%; 0 to 2.6 Sertoli-Leydig cell tumor 2 = 0.8%; 0.05 to 3.2 Peritoneal primary 1 = 0.4%; 0 to 2.6 Malignant mesothelioma 1 = 0.4%; 0 to 2.6 Other cancer 5 = 2.1%; 0.8 to 5.0 **Borderline tumors (LMP):** Serous epithelial 9 = 3.7%; 95% CI, 1.9 to 7.1 Mucinous epithelial 7 = 2.9%; 1.3 to 6.1 Endometrioid epithelial 1 = 0.4%; 0 to 2.6 Papillary mixed epithelial 1 = 0.4%; 0 to 2.6 **Benign:** Functional ovarian cyst 22 = 9.1%; 95% CI, 6.1 to 13.6 Serous cystadenoma 14 = 5.8%; 3.5 to 9.7 Mucinous cystadenoma 9 = 3.7%; 1.9 to 7.1 Brenner tumor 1 = 0.4%; 0 to 2.6 Endometrioma 10 = 4.1%; 2.2 to 7.6 Mature teratoma 6 = 2.5%; 1.1 to 5.5 Thecoma of fibroma 4 = 1.7%; 0.5 to 4.4 Tuboovarian abscess 4 = 1.7%; 0.5 to 4.4	**Comments:** --Clinical presentation not described --Patients scheduled for surgery; malignancy likely overrepresented **Quality assessment:** Size of population from which sample drawn: - (unclear) Number of cases: - (241) Patient selection: - (retrospective chart review) Application of reference standard: + (all had biopsy)

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				Hydrosalpinx 4 = 1.7%; 0.5 to 4.4 Paratubal cyst 1 = 0.4%; 0 to 2.6 Leiomyoma 22 = 9.1%; 6.1 to 13.6Pseudomyxoma 2 = 0.8%; 0.05 to 3.2 Endometriosis 1 = 0.4%; 0 to 2.6 Mesothelial cyst 1 = 0.4%; 0 to 2.6 Diverticular abscess 1 = 0.4%; 0 to 2.6	
Childers, Nasseri, and Surwit, 1996 **#6940**	**Geographical location:** Tucson, AZ **Dates:** 1991-1995 **Size of population:** 138 **Other** 138 with adnexal mass	**Age:** Mean: 52 Range: 9-91 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** All subjects had some concerning finding: CA-125 elevated: 39 (28%) Abnormal US: 127 (92%) Mass > 10 cm: 43 (32%)	**Benign: 119/138 (86.2%; 95% CI, 80.8 to 92.1)** 23 (16.7%; 11.6 to 24.2) cystadenoma 9 (6.5%) 3.4 to 12.3) mucinous cystadenoma 9 (6.5%; 3.4 to 12.3) cystadenofriboma 11 (8.0%; 4.5 to 14.1) benign teratoma 21 (15.2%; 10.3 to 22.5) endometrioma 2 (1.4%; 0.1 to 5.6) Brenner cell 1 (0.7%; 0 to 4.5) struma ovarii 9 (6.5%; 3.4 to 12.3) hydrosalpinx 3 (2.2%; 0.5 to 6.6) corpus luteum 6 (4.3%; 1.9 to 9.5) paraovarian cyst 15 (10.9%; 6.8 to 17.5) leiomyoma 6 (4.3%; 1.9 to 9.5) ovarian fibroma 3 (2.2%; 0.5 to 6.6) chronic tuboovarian abscess **Malignant: 19/138 (13.8%; 95% CI, 9.1 to 20.9) (16 of 19 adnexal primaries)** 5 (3.6%; 1.4 to 8.6) serous carcinoma 5 (3.6%; 1.4 to 8.6) serous cystadenocarcinoma 6 (4.3%; 1.9 to 9.5) endometroid carcinoma 3 (2.2%; 0.5 to 6.6) mixed endometroid and serous carcinoma Stage 1= 6 Stage 2 = 2 Stage 3 = 5 Unstaged = 3 (assumed to be Stage I), but 2 had recurrence	**Comments:** --Patients pre-selected for higher prevalence of malignancy -Clinical presentation not described **Quality assessment:** Size of population from which sample drawn: - (all women at one hospital) Number of cases: - (wide CIs) Patient selection: + (consecutive) Application of reference standard: + (all had biopsy)

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Cohen, Escobar, Scharm, et al., 2001 **#2460**	**Geographical location:** Chicago, IL **Dates:** Apr 1999-Jun 2000 **Size of population:** 71 **Other** Women with a complex pelvic mass undergoing laparotomy	**Age:** Range: 22-80 **Menopausal status (n [%]):** Pre (< 45): 40 (56%) Post (> 55): 31 (44%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Malignant: 13/71 = 18.3%; 95% CI, 11.4 to 29.7** Serous cystadenocarcinoma 5 = 7%; 2.9 to 16.2 Mucinous cystadenocarcinoma 1 = 1.4%; 0 to 8.6 Clear cell adenocarcinoma 1 = 1.4%; 0 to 8.6 Mixed mullerian 1 = 1.4%; 0 to 8.6 Malignant germ cell tumor 2 = 2.8%; 0.3 to 10.6 Metastatic sarcoma 1 = 1.4%; 0 to 8.6 Metastatic colon 2 = 2.8%; 0.3 to 10.6 **Borderline tumors (LMP): 1/71 = 1.4%; 0 to 8.6** Serous cystadenocarcinoma 1 = 1.4%; 0 to 8.6 **Benign: 57/71 = 80.3%; 71.9 to 89.7** Serous cystadenoma 9 = 12.7%; 7.0 to 23.2 Mucinous cystadenoma 6 = 8.4%; 3.9 to 18.0 Adenofibroma 10 = 14.1%; 8 to 24.8 Endometrioma 11 = 15.5%; 9.1 to 26.5 Cystic teratoma 13 = 18.3%; 11.4 to 29.7 Thecoma 1 = 1.4%; 0 to 8.6 Hydrosalpinx 4 = 5.6%; 2 to 14.4 Tamoxifen stimulation 2 = 2.8%; 0.3 to 10.6 Leiomyoma 1 = 1.4%; 0 to 8.6	**Comments:** --8/13 and the 1 borderline malignancy were in postmenopausal women --Clinical presentation not described --Patients scheduled for surgery; malignancy likely to be overrepresented **Quality assessment:** Size of population from which sample drawn: - (unclear) Number of cases: + Patient selection: - (only complex adnexal masses) Application of reference standard: + (all had biopsy)
DePriest,	**Geographical location:**	**Age:**	**Symptomatic (n [%]):**	Of the 6470 screened:	**Comments:**

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Gallion, Pavlik, et al., 1997 **#3650**	Lexington, KY **Dates:** Dec 1987-Dec 1993 **Size of population:** 6470; 8 found to have cancer (7 of these cancers were ovarian) **Screening study** Used TVUS in asymptomatic women > 50 or postmenopausal and women > 30 with positive family history of ovarian carcinoma	Mean: 58 Range: 30-92 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history of: Ovarian cancer: 24% Breast cancer: 30% Colon cancer: 15%	NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 99% (all but one was detected by imaging) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	7/6,470 malignant (ovarian) = 0.11%; 95% CI, 0.05 to 0.2 1/6,470 malignant (non-ovarian) = 0.02%; 0 to 0.1 83/6,470 benign = 1.2%; 1.0 to 0.12 **Malignant:** Granulosa cell tumor 3 = 0.05%; 95% CI, 0.01 to 0.14 Adenocarcinoma 2 = 0.03%; 0 to 0.12 Serous cystadenocarcinoma 1 = 0.02%; 0 to 0.1 Endometrioid carcinoma 1 = 0.02%; 0 to 0.1 Metastatic colon cancer 1 = 0.02%; 0 to 0.1 **Benign:** Serous cystadenoma 37 = 0.6%; 95% CI, 0.4 to 0.8 Endometriosis 18 = 0.3%; 0.2 to 0.4 Mucinous cystadenoma 3 = 0.05%; 0.01 to 0.14 Cystic teratoma 3 = 0.05%; 0.01 to 0.14 Hemorrhagic cyst 2 = 0.03%; 0 to 0.12 Fibroma/thecoma/Brenner tumor 4 = 0.06%; 0.02 to 0.2 Leiomyomata 4 = 0.06%; 0.02 to 0.2 Hydrosalpinx/paratubal 8 = 0.12%; 0.06 to 0.25 Other 4 = 0.06%; 0.02 to 0.2	--Overlap in data from previous study published by this group (DePriest, van Nagell Jr., Gallion, et al., 1993 [#6880]) --Most patients had either ovarian, breast or colon cancer family history **Quality assessment:** Size of population from which sample drawn: + (6,470/small city) Number of cases: - (8 with cancer) Patient selection: + (well-specified mix of postmenopausal women and high-risk younger women) Application of reference standard: + (all had biopsy)
DePriest, Shenson,	**Geographical location:** Lexington , KY	**Age:** Range: 3-74	**Symptomatic (n [%]):** NR	**Malignant: 13/121 = 10.7%; 95% CI, 6.5 to 17.9**	**Comments:** --Clinical presentation not described

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Fried, et al., 1993 **#6390**	**Dates:** Jan 1987-Jan1992 **Size of population:** 121 **Other** Women with ovarian mass undergoing laparotomy	**Menopausal status (n [%]):** Pre (< 45): 62 (51%) Post (> 55): 59 (49%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Serous cystadenocarcinoma 6 = 5%; 2.2 to 10.8 Mucinous cystadenocarcinoma 2 = 1.6%; 0.1 to 6.4 Granulosa cell tumor 1 = 0.8%; 0 to 5.1 Metastatic adenocarcinoma 2 = 1.6%; 0.1 to 6.4 Neurogenic sarcoma 1 = 0.8%; 0 to 5.1 Lymphoma 1 = 0.8%; 0 to 5.1 **Benign: 108/121 = 89.3%; 95% CI, 84 to 94.8** Serous cystadenoma 21 = 17.4%; 11.9 to 25.5 Mucinous cystadenoma 2 = 1.6%; 0.1 to 6.4 PID 18 = 14.9%; 9.8 to 22.7 Benign cysts = 17 = 14%; 9.1 to 21.8 Endometriosis 14 = 11.6%; 7.1 to 18.9 Hemorrhagic corpus luteum cyst 12 = 9.9%; 5.8 to 16.9 Teratoma 11 = 9.1%; 5.2 to 16 Fibroma = 5 = 4.1%; 1.6 to 9.7 Leiomyoma 4 = 3.3%; 1.1 to 8.6 Normal ovary 4 = 3.3%; 1.1 to 8.6	--Patients scheduled for surgery; malignancy likely overrepresented **Quality assessment:** Size of population from which sample drawn: - (not sure) Number of cases: - (wide CI) Patient selection: - (only those going to surgery) Application of reference standard: + (all had biopsy)
DePriest, van Nagell Jr., Gallion,	**Geographical location:** Lexington, KY	**Age:** Mean: 60 Range: 33-90	**Symptomatic (n [%]):** 0	**Benign: 41/3220 (1.3%; 95% CI, 0.9 to 1.7)** 21 (0.7%; 0.4 to 1.0) serous cystadenoma	**Comments:** --Majority, if not all, patients had either breast, ovarian, or colorectal

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
et al., 1993 **#6880**	**Dates:** Nov 1987-Jun 1992 **Size of population:** 3220 3 had cancer **Screening study (and most had a positive family history)**	**Menopausal status (n [%]):** Post (> 55): 100% **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history of: Ovarian cancer: 502 (15.6%) Breast cancer: 1034 (32.1%) Colorectal cancer: 678 (21.1%)	**Detected by exam (n [%]):** 0 **Detected by imaging (n [%]):** 44 had abnormal TVUS (1.4%) **Combination (n [%]):** 0 **Additional data used for diagnosis:** NR	4 (0.1%; 0.04 to 0.30) endometrioma 1 (0.03%; 0.0.2) cystadenofibroma 1(0.03%; 0.0.2) thecoma 1 (0.03%; 0.0.2) teratoma 2 (0.06%; 0 to 0.2) fibroma 3 (0.09%; 0.02 to 0.3) hydrodsalpinx 5 (0.16%; 0.06 to 0.4) paratubal cyst 3 (0.09%; 0.02 to 0.3) myoma **Malignant: 3/3220 (0.09%; 95% CI, 0.02, 0.29%)** 3 primary ovarian adenocarcinoma 2 Stage IA 1 Stage IIIB	cancer family history --True negative defined as negative biopsy or no diagnosed cancer within 1 year of ultrasound **Quality assessment:** Size of population from which sample drawn: - (unclear how representative – small city) Number of cases: + (although only 3 with cancer) Patient selection: - (some had family history) Application of reference standard: + (exploratory lab with biopsy)
Dottino, Levine, Ripley, et al., 1999 **#6920**	**Geographical location:** New York, NY **Dates:** Apr 1992-Apr 1996 **Size of population:** 160 **Other** Adnexal mass undergoing laparoscopic surgery	**Age:** Mean (SD): 52.2 (13.1) **Menopausal status (n [%]):** Pre (< 45): 75 (47%) Post (> 55): 85 (53%) **Race/ethnicity (n [%]):** White 146 (91%) **Risk factors (n [%]):** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Benign: 139/160 (86.9%; 95% CI, 81.9 to 92.2)** **Borderline: 8 (5%; 95% CI, 2.5 to 9.9)** **Malignant: 13 (8.1%; 95% CI, 4.8 to 13.7)** 6 epithelial cancers (3.7%; 1.6 to 8.2) 2 Stage 1A (1.2%; 0.09 to 4.8) 1 Stage 2C (0.6%; 0 to 3.9) 1 Stage 3A (0.6%; 0 to 3.9) 1 Stage 3C (0.6%; 0 to 3.9) 1 Stage 4 (0.6%; 0 to 3.9) 3 sex cord stromal tumors (1.9%; 0.4 to 5.7) 2 Sertoli-Leydig cell (1.2%; 0.09 to 4.8) 1 granulosa cell (0.6%; 0 to 3.9) 4 non-gynecologic cancers (4%; 0.8 to 6.6)	**Comments:** --Clinical presentation not described **Quality assessment:** Size of population from which sample drawn: + (large city) Number of cases: - (wide CIs) Patient selection: - (not described) Application of reference standard: + (all had biopsy)
Fleischer, Cullinan, Jones 3rd, et	**Geographical location:** Nashville, TN	**Age:** Mean: 50 Range: 17-88	**Symptomatic (n [%]):** NR	**Benign: 31/62 (50%; 95% CI, 39.5 to 63.5)** 10 (16.1%; 9.3 to 28.2) hemorrhagic corpus luteum	**Comments:** --Clinical presentation not described

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
al., 1996 **#3840**	**Dates:** 1990-1995 **Size of population:** 62 **Other** Patients who underwent Doppler for adnexal mass	**Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR Post (> 55): (over 50%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 100% **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1 (1.6%; 0 to 9.8) serous cyst 4 (6.4%; 2.3 to 16.4) cystadenomas 4 (6.4%; 2.3 to 16.4) endometriomas 8 (12.9%; 6.9 to 24.4) dermoid cysts 2 (3.2%; 0.4 to 12.1) ovarian fibroma 2 (3.2%; 0.4 to 12.1 leiomyoma **Malignant: 31/62 (50%; 95% CI, 39.5 to 63.5)** 16 (25.8%; 17.3 to 39.0) cystadeno-carcinomas 1 (1.6%; 0 to 9.8) papillary serous adenocarcinomas 1 (1.6%; 0 to 9.8) endometroid carcinoma 1 (1.6%; 0 to 9.8) dysgerminoma 4 (6.4%; 2.3 to 16.4) metastases 4 (6.4%; 2.3 to 16.4) germ cell tumors 4 (6.4%; 2.3 to 16.4) stromal tumors	**Quality assessment:** Size of population from which sample drawn: - (1 hospital) Number of cases: - (small and wide CIs) Patient selection: - (not described) Application of reference standard: + (all had biopsy)
Lin, Angel, DuBeshter, et al., 1993 **#4890**	**Geographical location:** Rochester NY **Dates:** Jun 1989-Jun 1990 **Size of population:** 80 **Other** Pelvic masses undergoing laparoscopic surgery	**Age:** Median: 56 Range: 19-88 **Menopausal status (n [%]):** Pre (< 45): NR Peri/Post: 62 (76%) **Race/ethnicity (n [%]):** White 72 (90%) Black 8 (10%) **Risk factors (n [%]):** Family history of: Ovarian/breast/colon cancer: 11 (14%)	**Symptomatic (n [%]):** 70 (87%) **Detected by exam (n [%]):** 80 (100%) --59 (74%) with discrete mass --21 (26%) ill-defined fullness **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Benign: 32/80 (40%; 95% CI, 30.9 to 52)** 23 (28.7%; 20.6 to 40.3) benign cyst 8 (10%; 5.2 to 19.2) other benign gynecologic condition 1 (1.2%; 0 to 7.7) diverticular disease **Borderline: 2/80 (2.5%; 95% CI, 0.3 to 9.5)** **Malignant: 46/80 (57.5%; 48 to 69.1)** 6 (7.5%; 3.4 to 16.1) colorectal carcinoma 1 (1.2%; 0 to 7.7) endometrial carcinoma 1 (1.2%; 0 to 7.7) vaginal carcinoma 2 (2.5%; 0.3 to 9.5) breast carcinoma 2 (2.5%; 0.3 to 9.5) lymphoma 4 (5%; 1.7 to 12.9) multiple sites 30 (37.5%; 28.6 to 49.5) ovarian carcinoma 6 (7.5%; 3.4 to 16.1) Stage 1 26 (32.5%; 24 to 44.3) Stage 3 4 (5%; 1.7 to 12.9) Stage 4	**Comments:** --Clinical presentation not described **Quality assessment:** Size of population from which sample drawn: - (1 hospital) Number of cases: - (wide CIs) Patient selection: - (not described) Application of reference standard: + (all had biopsy)
Modesitt, Pavlik, Ueland, et	**Geographical location:** Lexington, KY	**Age:** Range: 50-70+	**Symptomatic (n [%]):** 0	**Benign: 117/15106 (0.8%; 95% CI, 0.6 to 0.9)** 61 (0.4%; 0.3 to 0.5) serous cystadenomas	**Comments:** --Although cumulative incidence data are helpful, unable to calculate

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
al., 2003 **#5560**	**Dates:** 1987-2002 **Size of population:** 27 cancers/15,106 **Screening study with TVUS and followed up with Doppler and CA-125 if abnormal**	**Menopausal status (n [%]):** Post (> 55): 100% **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history: some but NR	**Detected by exam (n [%]):** 0 **Detected by imaging (n [%]):** 100% **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	14 (0.09%; 0.05 to 0.16) serous cystadenofibromas 9 (0.06%; 0.03 to 0.12) mucinous cystadenoma 9 (0.06%; 0.03 to 0.12) paraovarian 7 (0.05%; 0.02 to 0.1) fibrothecoma 5 (0.03%; 0.01 to 0.08) endometrioma 3 (0.02%; 0 to 0.06) cystic teratoma 1 (0.01%; 0 to 0.04) mucinous cystadenofibroma 8 (0.05%; 0.03 to 0.11) other **Malignant: 27/15106 (0.18%; 95% CI, 0.12 to 0.26)** 17 (0.11%; 0.07 to 0.18) Stage 1 4 (0.03%; 0.01 to 0.07) Stage 2 6 (0.04%; 0.02 to 0.09) Stage 3 Note: this is a separate group; of these 27, 10 had had simple ovarian cyst at one point in screening; 7 had additional morphologic abnormality, 2 had resolution of cyst before developing cancer, 1 had cancer in contralateral ovary Unilocular cyst – cumulative incidence by age: 50-54 1315/5229 (25.1%) 55-59 481/3278 (14.7%) 60-64 373/2694 (13.8%) 65-69 271/2008 (13.5%) 70+ 323/1897 (17.0%)	annual incidence rates given data provided **Quality assessment:** Size of population from which sample drawn: + (population-based) Number of cases: + (narrow CIs) Patient selection: Screening study, all over 50 Application of reference standard: + (subset of patient underwent biopsy)
Parker, Levine, Howard, et al., 1994	**Geographical location:** Santa Monica, Irvine, and Los Angeles, CA; Louisville, KY;	**Age:** Mean: 65 Range: 47-81	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):**	**All tumors were benign and were in postmenopausal women:** 27 (44.3%; 95% CI, 33.9 to 58.1) serous cystomas	**Comments:** --Clinical presentation not reported **Quality assessment:**

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
#910	Rochester, NY **Dates:** NR; published 1994 **Size of population:** 61 **Other** Laparoscopic management of benign-appearing cystic masses	**Menopausal status (n [%]):** Post (> 55): 61 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	NR **Detected by imaging (n [%]):** 100% **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	15 (24.6%; 16.2 to 37.8) serous cystadenomas 1 (1.6%; 0 to 9.9) mucinous cystadenomas 5 (8.2%; 3.5 to 18.7) cystadenofibromas 4 (6.6%; 2.4 to 16.7) hydrosalpinges 6 (9.8%; 4.6 to 20.8) paratubal cysts 3 (4.9%; 1.4 to 14.5) paraovarian cysts	Size of population from which sample drawn: + (multiple sites) Number of cases: - (no cancers; wide CIs) Patient selection: - (limited to benign-appearing cystic masses) Application of reference standard: + (all had biopsy)
Roman, Muder-spach, Stein, et al., 1997 **#6160**	**Geographical location:** Los Angeles, CA **Dates:** Jul 1992-Mar 1994 **Size of population:** 226 **Other:** Prospective study of women scheduled for removal of pelvic mass; **included women with pregnancy**	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 181 (80%) Post (> 55): 45 (20%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Of the 226 enrolled: 26/226 were malignant = 11.5%; 95% CI, 8.0 to 16.5 17/226 borderline tumors = 7.5%; 4.8 to 11.9 183/226 were benign = 81%; 76.1 to 86.2 **Malignant:** Epithelial cancer 15 = 6.6%; 95% CI, 4.1 to 10.9 Germ cell cancer 4 = 1.8%; 0.6 to 4.7 Stromal cancer 6 = 2.7%; 1.1 to 5.9 Sarcoma 1 = 0.4%; 0 to 2.8 **Borderline tumors:** LMP 17 = 7.5%; 95% CI, 4.8 to 11.9 **Benign:** Simple or functional cyst 46 = 20.4%; 95% CI, 15.8 to 26.3 Inflammatory process 18 = 8.0%; 5.1 to 12.4 Endometrioma 32 = 14.2%; 10.3 to 19.5 Cystic teratoma 32 = 14.2%; 10.3 to 19.5 Leiomyoma 11 = 4.9%; 2.7 to 8.7 Fibroma-thecoma 6 = 2.7%; 1.1 to 5.9 Cystadenoma 35 = 15.5%; 11.5 to 21 Cystadenofibroma 3 = 1.3%; 0.3 to 4.1	**Comments:** --Clinical presentation not described --Patients underwent surgery; malignancy likely overrepresented --Included women with pregnancy --Age range NR, but 80% reported to be premenopausal **Quality assessment:** Size of population from which sample drawn: - (unknown) Number of cases: - (226) Patient selection: + (prospectively collected information among women already scheduled for surgery) Application of reference standard: + (all had biopsy)
Schneider, Schneider, Reed, et al., 1993	**Geographical location:** Tucson, AZ **Dates:**	**Age:** Mean: 53 Median: 53 Range: 10-79	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):**	Of the 55 enrolled: 14/55 were malignant = 25.5%; 95% CI, 16.6 to 39.5 2/55 were borderline tumors = 3.6%; 0.5	Comments: --Clinical presentation not described --Patients underwent surgery and therefore malignancy likely to be

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	NR; published 1993		NR	to13.5 39/55 were benign 70.9%; 60.5 to 83.4	over-represented
#4830	**Size of population:** 55 **Other:** Patients undergoing surgery for adnexal mass	**Menopausal status (n [%]):** Pre (< 45): 22 (40%) Post (> 55): 33 (60%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Malignant:** Endometriod cancer 6 = 10.9%; 95% CI, 5.2 to 22.9 Adenocarcinoma, undifferentiated 3 = 5.5%; 1.5 to 16 Carcinosarcoma 1 = 1.8%; 0 to 11 Clear-cell adenocarcinoma 1 = 1.8%; 0 to 11 Krukenberg tumor (primary gastric) 1 = 1.8%; 0 to 11 Malignant Brenner tumor 1 = 1.8%; 0 to 11 Leiomyosarcoma 1 = 1.8%; 0 to 11 **Borderline tumors (LMP):** Serous cystadenocarcinoma 2 = 3.6%; 95% CI, 0.5 to13.5 **Benign:** Serous cystadenoma 12 = 21.8%; 95% CI, 13.6 to 35.6 Endometriosis 4 = 7.3%; 2.7 to 18.4 Mucinous cystadenoma 3 = 5.5%; 1.5 to 16 Follicular cyst 3 = 5.5%; 1.5 to 16 Adenofibroma 1 = 1.8%; 0 to 11 Brenner tumor 1 = 1.8%; 0 to 11 Corpus luteum cyst 1 = 1.8%; 0 to 11 Fibroma 1 = 1.8%; 0 to 11 Mature teratoma 1 = 1.8%; 0 to 11 Mixed stromal cell tumor 1 = 1.8%; 0 to 11 Serous Cystadenofibroma 1 = 1.8%; 0 to 11 Leiomyoma 4 = 7.3%; 2.7 to 18.4 Paraovarian cyst 3 = 5.5%; 1.5 to 16 Hydrosalpinx 2 = 3.6%; 0.5 to 13.5 Peritoneal inclusion cyst 1 = 1.8%; 0 to 11	**Quality assessment:** Size of population from which sample drawn: - (unknown) Number of cases: - (55) Patient selection: - (cross-sectional) Application of reference standard: + (all had biopsy)
Scoutt, McCarthy, Lange, et al., 1994	**Geographical location:** Connecticut **Dates:** 1988-1990	**Age:** Median: 40 Range: 2-87 **Menopausal status**	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR	**Benign: 87/109 (79.8%; 95% CI, 72.8 to 87.5)** 17 (15.6%; 10.2 to 24.1) leiomyoma 19 (17.4%; 11.7 to 26.1) dermoid 13 (11.9%; 7.2 to 19.8) endometrioma	**Comments:** --Clinical presentation not described **Quality assessment:** Size of population from which

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
#4530	**Size of population:** 109 masses with MRI **Other** Clinical masses that underwent MRI and then biopsy	**(n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Detected by imaging (n [%]):** NR **Combination (n [%]):** 109 (100%) **Additional data used for diagnosis:** NR	9 (8.3%; 4.4 to 15.4) hemorrhagic cysts 9 (8.3%; 4.4 to 15.4) simple cysts 5 (4.6%; 1.8 to 10.8) serous cystadenoma 3 (2.7%; 0.7 to 8.3) mucinous cystadenoma 3 (2.7%; 0.7 to 8.3) friboma 3 (2.7%; 0.7 to 8.3) tuboovarian abscess 3 (2.7%; 0.7 to 8.3) paratubal cyst 1 (0.9%; 0 to 5.7) fibrothecoma 1 (0.9%; 0 to 5.7) leutinized thecoma 1 (0.9%; 0 to 5.7) hematosalpinx **Malignant: 22/109 (20.2%; 95% CI, 14 to 29.2)** 5 (4.6%; 1.8 to 10.8) papillary serous cystadenocarcinoma 4 (3.7; 1.2 to 9.6%) metastatic adenocarcinoma 3 (2.7%; 0.7 to 8.3) mucinous cystadenocarcinoma 3 (2.7%; 0.7 to 8.3) endometroid carcinoma 2 (1.8%; 0.2 to 7.0) adenocarcinoma 1 (0.9%; 0 to 5.7) immature teratoma 1 (0.9%; 0 to 5.7) embryonal cell carcinoma 1 (0.9%; 0 to 5.7) dysgerminoma 1 (0.9%; 0 to 5.7) granulosa cell tumor 1 (0.9%; 0 to 5.7) endometrial carcinoma	sample drawn: - (1 hospital) Number of cases: - (wide CIs) Patient selection: - (suspected mass who had MRI) Application of reference standard: + (all had biopsy)
Shen-Gunther and Mannel, 2002 **#2090**	**Geographical location:** Las Vegas, NV; Oklahoma City, OK **Dates:** Jan 1994-Dec 1994 and	**Age:** Median: 58 Range: 18-86 **Menopausal status (n [%]):**	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 6%	**Benign: 57/125 (45.6%; 95% CI, 37.8 to 55.1)** 22 (17.6%; 12.1 to 25.6) serous cystadenoma 3 (2.4%; 0.6 to 7.3) mucinous cystadenoma 4 (3.2%; 1.1 to 8.4) friboma	**Comments:** --Large proportion of subjects had ascites on exam or imaging – very high prevalence of malignancy **Quality assessment:**

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Jan 1999-Dec 2001 **Size of population:** 125 **Other** Patients treated for pelvic mass	NR **Race/ethnicity (n [%]):** White 82% Black 9% Hispanic 2% Asian 4% American Indian 3% **Risk factors (n [%]):** NR	**Detected by imaging (n [%]):** Ultrasound 46% CT 18% Both 31% **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	2 (1.6%; 0.1 to 6.2) thecoma 5 (4%; 1.6 to 9.4) teratoma 5 (4%; 1.6 to 9.4 follicular cyst 1 (0.8%; 0 to 5) paratubal cyst 5 (4%; 1.6 to 9.4 hemorrhagic cysts 2 (1.6%; 0.1 to 6.2) tuboovarian adhesions 8 (6.4%; 3.2 to 12.5) endometrioma **Borderline (LMP): 12/125 (9.6%; 95% CI, 5.6 to 16.4)** 8 (6.4%; 3.2 to 12.5) serous low malignant potential 4 (3.2%; 1.1 to 8.4) mucinous low malignant potential **Malignant: 56/125 (44.8%; 37.1 to 54.3)** 39 (31.2%; 24.2 to 40.3) serous cystadenocarcinoma 2 (1.6%; 0.1 to 6.2) mucinous cystadenocarcinoma 4 (3.2%; 1.1 to 8.4) endometroid carcinoma 4 (3.2%; 1.1 to 8.4) primary peritoneal carcinoma 2 (1.6%; 0.1 to 6.2) clear cell carcinoma 1 (0.8%; 0 to 5) undifferentiated adenocarcinoma 2 (1.6%; 0 to 5) immature teratoma Stage 1 = 11 Stage 2 = 1 Stage 3 = 33 Stage 4 = 5 Unstaged = 6	Size of population from which sample drawn: + (2 cities) Number of cases: - (wide CIs) Patient selection: - (2 separate time frames introduces bias; also high prevalence of ascites) Application of reference standard: + (all had biopsy)
Smikle, Lunt, and Hankins, 1995 **#6290**	**Geographical location:** San Antonio, TX **Dates:** Jun 1990-Aug 1992 **Size of population:**	**Age:** Range: < 20 and > 61 **Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging**	**Benign: 169/195 (86.7%; 95% CI, 82.1 to 95.0)** 37 (19.0%; 14.3 to 25.3) serous cystadenoma 11 (5.6%; 3.2 to 10.1) mucinous cystadenoma 26 (13.3%; 9.4 to 19.1) hemorrhagic cysts	**Comments:** --Clinical presentation not described **Quality assessment:** Size of population from which sample drawn: - (military hospital) Number of cases: - (26 cancers and

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	195 **Other** Surgical cases with preoperative diagnosis of pelvic mass	Post (> 55): 78 (40%) 51 and older **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**(n [%]):** NR **Combination (n [%]):** 195 (100%) **Additional data used for diagnosis:** NR	20 (10.3%; 6.8 to 15.6) endometriosis 17 (8.7%; 5.5 to 13.8) teratoma (mature) 7 (3.6%; 1.7 to 7.6) cyst of Morgagni Data not provided for 27 cases **Malignant: 26/195 (13.3%; 95% CI, 9.4 to 19.1)** 14 (7.2%; 4.3 to 11.9) serous cystadenocarcinoma 5 (2.6%; 1.0 to 6.3) mucinous cystadenocarcinoma 1 (0.5%; 0 to 3.2) endometroid carcinoma 4 (2.1%; 0.7 to 5.4) undifferentiated adenocarcinoma 1 (0.6%; 0 to 3.2) granulosa cell carcinoma **Benign mass by age:** Age ≤ 50 (n = 117) Serous cystadenoma: 19 (16.2%; 95% CI, 10.9 to 24.4) Functional cyst: 20 (17.1%; 11.6 to 25.4) Hydrosalpinx/tuboovarian abscess: 18 (15.4%; 10.1 to 23.5%) Endometriosis: 16 (13.7%; 8.7 to 21.5) Mature teratoma: 11 (9.4%; 5.4 to 16.5) Mucinous cystadenoma: 3 (2.6%; 0.6 to 7.8) Cyst of Morgagni: 4 (3.4%; 1.1 to 8.9) Age > 50 (n = 78) Serous cystadenoma: 18 (23.1%;15.6 to 34.4) Functional cyst: 6 (7.7%; 3.5 to 16.5) Hydrosalpinx/ tuboovarian abscess: 5 (6.4%; 2.6 to 14.9) Endometriosis: 4 (5.1%; 1.8 to13.2) Mature teratoma: 6 (7.7%; 3.5 to 16.5) Mucinous cystadenoma: 8 (10.3%; 5.3 to 19.7) Cyst of Morgagni: 3 (3.9%; 1.0 to 11.5%)	wide CIs) Patient selection: - (all surgical cases) Application of reference standard: + (all had biopsy)
Troiano, Quedens-Case, and Taylor, 1997	**Geographical location:** New Haven, CT **Dates:** 1991-1996	**Age:** Mean: Approx. 45 Range: 18-79 **Menopausal status**	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 100% suspected mass on	**Malignant: 17/144 = 11.8%; 95% CI, 7.6 to 18.4** Serous cystadenocarcinoma 7 = 4.9%; 2.3 to 10 Mucinous cystadenocarcinoma 1 = 0.7%; 0	**Comments:** --Not all subjects went to surgery; better generalizability, but possible error in diagnosis

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
#3680	**Size of population:** 144 patients **Other** Patients with suspected mass on exam and referred for US; not all went on to surgery, but all had followup	**(n [%]):** Pre (< 45): 101 (70%) Post (> 55): 42 (29%) Missing 1 case **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	exam **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	to 4.3 Endometrioid carcinoma 1 = 0.7%; 0 to 4.3 Embryonal cell carcinoma 1 = 0.7%; 0 to 4.3 Granulosa cell 1 = 0.7%; 0 to 4.3 Fallopian tube carcinoma 1 = 0.7%; 0 to 4.3 Endometrial 1 = 0.7%; 0 to 4.3 Metastatic 4 = 2.8%; 0.9 to 7.3 Borderline tumors 3/144 = 2.1%; 0.5 to 6.3 Borderline papillary serous 3 = 2.1%; 0.5 to 6.3 **Benign: 97/144 = 67.4%; 95% CI, 60.3 to 75.3** Serous cystadenoma 1 = 0.7%; 0 to 4.3 Mucinous cystadenoma 3 = 2.1%; 0.5 to 6.3 Functional ovarian cyst 3 = 2.1%; 0.5 to 6.3 Paratubal cyst 4 = 2.8%; 0.9 to 7.3 Ovarian dermoid cyst 4 = 2.8%; 0.9 to 7.3 Fibroma or thecoma 2 = 1.4%; 0.1 to 5.4 Cystadenofibroma 1 = 0.7%; 0 to 4.3 Endometriosis or hemorrhagic cyst 16 = 11.1%; 7 to 17.6 Leiomyomas or adenomyosis 43 29.9%; 23.4 to 38.3 Leiomyomas with endometriosis 6 = 4.2%; 1.8 to 9.1 Leiomyomas with simple ovarian cyst 2 = 1.4%; 0.1 to 5.4 Leiomyomas with paratubal cyst 1 = 0.7%; 0 to 4.3 Leiomyomas with ovarian fibroma 1 = 0.7%; 0 to 4.3 Leiomyomas with Brenner tumor 1 = 0.7%; 0 to 4.3 Leiomyomas with ovarian cystadenoma 2 = 1.4%; 0.1 to 5.4 Cirrhosis 1 = 0.7%; 0 to 4.3 Pregnancy 1 = 0.7%; 0 to 4.3 **No biopsy because ultrasound negative: 27 = 18.7%; 95% CI, 13.4 to 26.3**	**Quality assessment:** Size of population from which sample drawn: - (not clear) Number of cases: - (wide CIs) Patient selection: + (better than the others – all with suspected mass on exam) Application of reference standard: + (not all had biopsy, but all had followup)
Twickler, Forte, Santos-	**Geographical location:** Dallas, TX	**Age:** Mean: 38.6 Range: 15-80	**Symptomatic (n [%]):** NR	**Malignant: 14/244 = 5.7%; 95% CI, 3.4 to 9.6** Serous 4 = 1.6%; 0.5 to 4.4	**Comments:** --Not all subjects went to surgery; better generalizability, but more

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Ramos, et al., 1999 **#3080**	**Dates:** Feb 1993-Aug 1996 **Size of population:** 244 women **Other** 304 had ultrasound for mass, and 217 had surgery and another 27 had ultrasound followup, for a total of 244	**Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR	**Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Metastasis 4 = 1.6%; 0.5 to 4.4 Adenocarcinoma 2 = 0.8%; 0.05 to 3.2 Mucinous 1 = 0.4%; 0 to 2.6 Small cell 1 = 0.4%; 0 to 2.6 Sarcoma 1 = 0.4%; 0 to 2.6 Mixed germ cell 1 = 0.4%; 0 to 2.6 **Borderline tumors: 16/244 = 6.6%; 95% CI, 4.1 to 10.6** Mucinous 8 = 3.3%; 1.6 to 6.5 Serous 5 = 2%; 0.8 to 4.9 Granulosa cell 2 = 0.8%; 0.05 to 3.2 Endometrioid 1 = 0.4%; 0 to 2.6 **Benign: 214/244 = 87.7%; 83.7 to 91.9** Simple functional cyst 69 = 28.3%; 23.2 to 34.5 PID mass 25 = 10.2%; 7.1 to 14.9 Endometriomas 13 = 5.3%; 3.1 to 9.1 No ovarian mass 7 = 2.9%; 1.3 to 6 Non-defined ovarian cystic disease 8 = 3.3%; 1.6 to 6.5 Para-ovarian cyst 1 = 0.4%; 0 to 2.6 Paratubal cyst 1 = 0.4%; 0 to 2.6 Fibrovascular ampullary mass 1 = 0.4%; 0 to 2.6 Ectopic mass 1 = 0.4%; 0 to 2.6 Ovarian lymphocele 1 = 0.4%; 0 to 2.6 Peritoneal cyst 1 = 0.4%; 0 to 2.6 Mesonephric cyst 1 = 0.4%; 0 to 2.6 Dermoid/cystic teratoma 35 = 14.3%; 10.6 to 19.5 Serous cystadenoma 19 = 7.8%; 5.1 to 12 Cystadenofibroma 13 = 5.3%; 3.1 to 0.1 Mucinous cystadenoma 7 = 2.9%; 1.3 to 6 Cystadenoma (unspecified) 3 = 1.2%; 0.3 to 3.8 Fibroma 2 = 0.8%; 0.05 to 3.2 Fibrothecoma 1 = 0.4%; 0 to 2.6 Seromucinous 1 = 0.4%; 0 to 2.6 Other 4 = 1.6%; 0.5 to 4.4	possibility of error **Quality assessment:** Size of population from which sample drawn: - (not known) Number of cases: + Patient selection: - (not clear) Application of reference standard: - (not all had biopsy and some were lost to followup)
van Nagell Jr., DePriest, Reedy, et	**Geographical location:** Kentucky **Dates:**	**Age:** Mean (SD): 54.7 (10.7) Range: 25-92	**Symptomatic (n [%]):** 0 **Detected by exam (n [%]):**	**Benign: 155/14,469 (1.1%; 95% CI, 0.9 to 1.2)** 78 (0.5%; 0.4 to 0.7) serous cystadenomas 25 (0.2%; 0.1 to 0.3) endometriomas	**Quality assessment:** Size of population from which sample drawn: + (screening study) Number of cases: + (large screening

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
al., 2000 **#2730**	1987-1999 **Size of population:** 17 cancers/14,469 (actually 3 were borderline) **Screening study (most had positive family history; 180 had a biopsy)**	**Menopausal status (n [%]):** All ≥ 50 or ≥ 25 with family history **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history of: Ovarian cancer: 23% Breast cancer: 34% Colon cancer: 23%	3/17 cancers were palpated on exam but not detected on exam **Detected by imaging (n [%]):** 100% had TVUS **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	10 (0.07%; 0.04 to 0.1) mucinuos cystadenomas 11 (0.08%; 0.04 to 0.1) cystic teratomas 13 (0.09%; 0.05 to 0.2) fibroma/thecoma 4 (0.03%; 0.01 to 0.07) leiomyoma 14 (0.1%; 0.06 to 0.2) hydrosaplinx/ paratubal cyst **Borderline (LMP): 3/14,469 (0.02%; 95% CI, 0 to 0.06)** All 3 serous low malignant potential **Malignant: 14/14469 (0.1%; 95% CI, 0.06 to 0.2)** 1 (0.01%; 0 to 0.04) serous cystadenocarcinoma 1 (0.01%; 0 to 0.04) mucinous cystadenocarcinoma 3 (0.02%; 0 to 0.06) endometroid carcinoma 6 (0.04%; 0.02 to 0.09) undifferentiated adenocarcinoma 3 (0.02%; 0 to 0.06) granulosa cell carcinoma	study with narrow CIs) Patient selection: - (most with family history) Application of reference standard: + (all with abnormal TVUS had biopsy)
Vasilev, Schlaerth, Campeau, et al., 1988 **#6770**	**Geographical location:** Los Angeles, CA **Dates:** Apr 1984-Feb 1986 **Size of population:** 182 non-consecutive patients with pelvic mass **Other**	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):**	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):**	**Malignant: 15/182 = 8.2%; 95% CI, 5.1 to 13.4** Serous cystadenocarcinoma 4 = 2.2%; 0.7 to 5.8 Mucinous cystadenocarcinoma 2 = 1.1%; 0.1 to 4.3 Endometrioid carcinoma 1 = 0.5%; 0 to 3.4 Leiomyosarcoma1 = 0.5%; 0 to 3.4 Gastric Krukenberg tumor 1 = 0.5%; 0 to 3.4 Hypernephroma 1 = 0.5%; 0 to 3.4	**Comments:** --8 of 10 masses in women over 50 were malignant --Selection criteria for inclusion in series not included **Quality assessment:** Size of population from which sample drawn: - Number of cases: - Patient selection: - (all had mass)

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Non-consecutive series of patients scheduled for surgery for adnexal mass	NR	NR **Additional data used for diagnosis:** NR	Lymphoma 1 = 0.5%; 0 to 3.4 Melanoma 1 = 0.5%; 0 to 3.4 Uterine leiomyosarcoma 3 = 1.6%; 0.4 to 5 **Borderline tumors 3/182 = 1.6%; 95% CI, 0.4 to 5** Serous low malignant potential 1 = 0.5%; 0 to 3.4 Mucinous low malignant potential 2 = 1.1%; 0.1 to 4.3 **Benign: 164/182 = 90.1%; 95 % CI, 85.9 to 94.5** Adhesions complex 2 = 1.1%; 0.1 to 4.3 Paratubal cysts 6 = 3.3%; 1.4 to 7.3 Ectopic pregnancy 2 = 1.1%; 0.1 to 4.3 Acute salpingitis 12 = 6.6%; 3.8 to 11.4 Chronic salpingitis 3 = 1.6%; 0.4 to 5 Serous cystadenoma 9 = 4.9%; 2.6 to 9.4 Mucinous cystadenoma 4 = 2.2%; 0.7 to 5.8 Benign cystic teratoma 13 = 7.1%; 4.2 to 12.1 Fibroma 2 = 1.1%; 0.1 to 4.3 Brenner tumor 1 = 0.5%; 0 to 3.4 Endometrioma 5 = 2.7%; 1.1 to 6.5 Simple ovarian cyst 2 = 1.1%; 0.1 to 4.3 Leiomyoma 71 = 39%; 32.6 to 46.7 adenomyosis 9 = 4.9%; 2.6 to 9.4 Leiomyomas with endometriosis 2 = 1.1%; 0.1 to 4.3 Leiomyomas with adenomyosis 8 = 4.4%; 2.2 to 8.7 Leiomyomas with chronic salpingitis 5 = 2.7%; 1.1 to 6.5 Leiomyomas with endometriosis and adenomyosis 1 = 0.5%; 0 to 3.4 Leiomyomas with Brenner tumor and mucinous cystadenoma 1 = 0.5%; 0 to 3.4 Leiomyomas with serous cystadenoma 1 = 0.5%; 0 to 3.4 Leiomyomas with adenomyosis and chronic salpingitis 2 = 1.1%; 0.1 to 4.3 Leiomyomas with endometriosis and chronic salpingitis 1 = 0.5%; 0 to 3.4	Application of reference standard: - (not all had biopsy)

Evidence Table 1 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				Leiomyomas and salpingitis and paratubal cyst 1 = 0.5%; 0 to 3.4 Cystadenofibroma and Leiomyoma and endometriosis 1 = 0.5%; 0 to 3.4	

Evidence Table 2: *Question 2: What are the sensitivity, specificity, and reliability of the bimanual pelvic examination?*

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Adonakis, Para-skevaidis, Tsiga, et al., 1996 **#810**	**Geographical location:** Greece **Dates:** Mar 1991-Jun 1993 **Size of population:** 2000 **Screening study** **Reference standard:** US if BME abnormal or ambiguous; surgery if US positive; 12-month CA-125 if negative **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No; performed by 3 gynecological oncologists **Statistical tests used:** Se, Sp **Blinding:** Yes **Definition of positive and negative on screening test:** Positive exam: palpable adnexal mass Ambiguous: origin of mass unclear or inadequate exam	**Age:** Mean (SD): 58.1 (6.9) Range: 45-80 **Menopausal status (n [%]):** Pre: 405 (20%) Peri: 293 (15%) Post: 1302 (65%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women > 45, no evidence of adnexal pathology, agreed to participate **Exclusion criteria:** History of ovarian cancer or any other malignancy, history of bilateral oophorectomy, ascites	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** 50 (3%) positive exam 115 (6%) "ambiguous" exam **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Women with elevated CA-125 or and abnormal or ambiguous BME were recalled for TVUS. Only women with +TVUS were referred for further management	Screening study	1) Benign vs. malignant: Dis+ / Dis- / Tot T+: 2 / 57 / 59 T-: 1 / 1940 / 1941 Tot: 3 / 1997 / 2000 Value (Lower 95% CI – Upper 95% CI) Se: 66.7% (13.3% – 100.0%) Sp: 97.1% (96.4% – 97.9%) PPV: 3.4% (0.0% – 8.0%) NPV: 99.9% (99.8% – 100.0%)	**Comments:** --1 tumor LMP grouped in with 2 other malignancies --"Ambiguous" BME was classed as Test -, although all patients with ambiguous BME had TVUS to further evaluate --Borderline tumors considered Dis+ **Quality assessment:** Reference standard: + (followup with CA-125 at 12 months reasonable for screening study) Verification bias: + (all test negatives had 12-month CA-125) Test reliability/variability: - Sample size: + Statistical tests: + Blinding: + Definition of +/- on screening test: - ("palpable" not precise)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Andolf, Jorgensen, and Astedt, 1990 **#1200**	**Geographical location:** Sweden **Dates:** Oct 1984-Jul 1987 **Size of population:** 801 **Screening study** For women at high risk for ovarian cancer **Reference standard:** All had US and some had biopsies **Reference standard applied to all test negatives?:** No **Test reliability established?:** No **Statistical tests used:** None **Blinding:** NR **Definition of positive and negative on screening test:** NR	**Age:** Range: 40-70 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history: 190 (23.7%) **Inclusion criteria:** Women older than 40 with either abdominal pain, nulliparity, family history of breast, ovarian, or endometrial cancer, or previous history of cancer **Exclusion criteria:** NR	**Symptomatic (n [%]):** 419 (52.3%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	One gynecologist clinical examiner, and then a midwife did the US	1) Abnormal vs. normal US: Dis+ / Dis- / Tot T+: 55 / 51 / 106 T-: 108 / 587 / 695 Tot: 163 / 638 / 801 Value (Lower 95% CI, Upper 95% CI) Se: 33.7% (26.5%, 41.0%) Sp: 92.0% (89.9%, 94.1%) PPV: 51.9% (42.4%, 61.4%) NPV: 84.5% (81.8%, 87.2%)	**Comments:** --US by midwife and not a radiologist; only 30 abnormal scans went on to surgery --2 endometrial carcinomas and 1 borderline ovarian tumor **Quality assessment:** Reference standard: - (all had US and not all had biopsy) Verification bias: - Test reliability/variability: - Sample size: - (no ovarian cancer) Statistical tests: - Blinding: - (not stated) Definition of +/- on screening test: -

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Balbi, Musone, Menditto, et al., 2001 **#2320**	**Geographical location:** Naples, Italy **Dates:** Jan 1996-Mar 2000 **Size of population:** 92 **Case series** **Reference standard:** Histopathological diagnosis **Reference standard applied to all test negatives?:** No – 18 women with "clearly benign" masses not verified; 2 patients with "clearly malignant" disease (metastases) also excluded **Test reliability established?:** Not for PE or CA-72-4 Uncertain for US and RI Yes for CA-125 **Statistical tests used:** Se, Sp, multivariate logistic analysis **Blinding:** NR **Definition of positive and negative on screening test:** PE: "malignant clinical impression"	**Age:** Range: 40-80 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** "Women with a pelvic mass originating in the ovary" **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Women with pelvic mass prior to surgery. Physical exam by standard protocol. Examiner was asked to predict benign or malignant.	1) Benign vs. malignant: Dis+ / Dis- / Tot T+: 20 / 13 / 33 T-: 2 / 37 / 39 Tot: 22 / 50 / 72 Value (Lower 95% CI – Upper 95% CI) Se: 90.0% (77.5% – 100.0%) Sp: 74.0% (61.8% – 86.2%) PPV: 60.6% (43.9% – 77.3%) NPV: 94.9% (87.9% – 100.0%)	**Comments:** --20 patients excluded for reasons that seem to indicate there wasn't blinding --Vague definition of PE --Although RI measured, not included in definition of +US **Quality assessment:** Reference standard: + Verification bias: - ("clearly benign" excluded) Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: - (definition of "clinical impression" not provided)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Buckshee, Temsu, Bhatla, et al., 1998 **#710**	**Geographical location:** India **Dates:** May 1995-Apr 1997 **Size of population:** 34 non-consecutive women with 36 tumors **Other:** Women scheduled for surgery for adnexal mass **Reference standard:** Biopsy **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** McNemar test **Blinding:** Yes **Definition of positive and negative on screening test:** Yes Clinical diagnosis-benign vs. malignant	**Age:** Range: 20 to > 50 **Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR Post (> 55): 5 > 50 **Race/ethnicity (n [%]):** Indian **Risk factors (n [%]):** Family history: 1 (3%) **Inclusion criteria:** Scheduled for surgery **Exclusion criteria:** Known cancer and with evidence of extensive/metastatic disease on US or CT scan	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 10 (2.8%) **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	One gynecologist clinical examiner	1) Malignant vs. benign: Dis+ / Dis- / Tot T+: 7 / 3 / 10 T-: 2 / 24 / 26 Tot: 9 / 27 / 36 Value (Lower 95% CI – Upper 95% CI) Se: 77.8% (50.6% – 100.0%) Sp: 88.9% (77.0% – 100.0%) PPV: 70.0% (41.6% – 98.4%) NPV: 92.3% (82.1% – 100.0%)	**Quality assessment:** Reference standard: + (all had biopsy) Verification bias: + Test reliability/variability: - Sample size: - (small study) Statistical tests: + Blinding: + (yes) Definition of +/- on screening test: - (not very specific; essentially a clinical impression of benign vs. malignant)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Dowd, Quinn, Rome, and Koh, 199 **#4680**	**Geographical location:** Melbourne, Australia **Dates:** 1978-1989 **Size of population:** 264 (n = 225 with definite clinical impression) **Case series** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced or measured **Statistical tests used:** Se, Sp, NPV, PPV **Blinding:** No **Definition of positive and negative on screening test:** "Mass described as 'hard, irregular, fixed, attached to other structures', or associated with ascites, or a specific statement from a consultant gynaecologist of the suspected malignant nature of the mass"	**Age:** Mean: 47 Range: 15-89 **Menopausal status (n [%]):** Pre (< 45): 78 (61%) Peri (45-55): Post (> 55): 50 (39%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Suspected pelvic mass, CA-125 level available **Exclusion criteria:** Screening, inadequate documentation of clinical findings of pathology	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Not described; presumably, in outpatient setting	1) Benign vs. malignant - all patients: Dis+ / Dis- / Tot T+: 56 / 15 / 71 T-: 54 / 100 / 154 Tot: 110 / 115 / 225 Value (95% CI) Se: 51.0% (41.7% - 60.3%) Sp: 87.0% (80.8% - 93.2%) PPV: 79.0% (69.5% - 88.5%) NPV: 65.0% (57.5% - 72.5%) 2) Other: Values also reported by menopausal status, but insufficient data to construct 2x2 table: Premenopausal: Sensitivity 31%, specificity 95%, PPV 76%, NPV 75% Postmenopausal: Sensitivity 59%, specificity 75%, PPV 79%, NPV 54%	**Comments:** --Examiners not blinded to history, possibly other findings --High prevalence of malignancy --History not provided; unclear how many subjects were symptomatic **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: + (but confidence intervals not given) Statistical tests: + Blinding: - Definition of +/- on screening test: - (definitions not explicit)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Finkler, Benacerraf, Lavin, et al., 1988 **#1230**	**Geographical location:** Boston, MA **Dates:** Nov 1986-Apr 1987 **Size of population:** 106 **Other:** Consecutive patients with adnexal mass scheduled for surgery **Reference standard:** Biopsy **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Fisher's exact test **Blinding:** NR **Definition of positive and negative on screening test:** Clinicians asked to judge the clinical appearance of the mass based on history and physical exam combined	**Age:** Mean: 45.2 Range: 17-84 **Menopausal status (n [%]):** Pre (< 45): 74 (69.8%) Post (> 55): 32 (30.2%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Adnexal mass scheduled for surgery **Exclusion criteria:** US unavailable or uninterpretable; pregnancy; known cancer	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	One gynecologist examiner who gave his/her verbal clinical impression before surgery	1) Total study - clinical impression is test, and malignant (yes/no) is disease state: Dis+ Dis- Tot T+ 16 6 22 T- 21 59 80 Tot 37 65 102 Value / Lower 95% CI / Upper 95% CI Se 43.2% 27.3% 59.2% Sp 90.8% 83.7% 97.8% PPV 72.7% 54.1% 91.3% NPV 73.8% 64.1% 83.4% 2) Premenopausal group: Dis+ Dis- Tot T+ 3 4 7 T- 15 48 63 Tot 18 52 70 Value / Lower 95% CI / Upper 95% CI Se 16.7% 0.0% 33.9% Sp 92.3% 85.1% 99.6% PPV 42.9% 6.2% 79.5% NPV 76.2% 65.7% 86.7% 3) Postmenopausal group: Dis+ Dis- Tot T+ 13 2 15 T- 6 11 17 Tot 19 13 32 Value / Lower 95% CI / Upper 95% CI Se 68.4% 47.5% 89.3% Sp 84.6% 65.0% 100.0% PPV 86.7% 69.5% 100.0% NPV 64.7% 42.0% 87.4%	**Comments:** --Definition of a positive physical examination is the "impression of clinical exam" that includes history **Quality assessment:** Reference standard: + (all had biopsy) Verification bias: - Test reliability/variability: - Sample size: - (small study) Statistical tests: - Blinding: - (not stated) Definition of +/- on screening test: - ("impression of clinical exam" that includes history)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Grover and Quinn, 1995 **#830**	**Geographical location:** Melbourne, Australia **Dates:** NR **Size of population:** 2623 **Screening study** Healthy volunteers **Reference standard:** US if mass or elevated CA-125 Surgery or 12-month followup questionnaire **Reference standard applied to all test negatives?:** Yes – 12-month followup questionnaire for all **Test reliability established?:** No; single examiner **Statistical tests used:** Chi-square **Blinding:** No **Definition of positive and negative on screening test:** Premenopausal: "larger than normal" Postmenopausal: palpable	**Age:** Mean: 51 Range: 25-92 **Menopausal status (n [%]):** Pre (< 45): 1121 (43%) Peri (45-55): 384 (15%) Post (> 55): 1118 (42%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Asymptomatic, recruited (not clear how) **Exclusion criteria:** NR	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Single examiner	1) All women: Dis+ Dis- Tot T+ 0 40 40 T- 1 2582 2583 Tot 1 2622 2623 Value / Lower 95% CI / Upper 95% CI Se 0.0% 0.0% 100.0% Sp 98.5% 98.0% 98.9% PPV 0.0% 0.0% 7.5% NPV 100.0% 99.9% 100.0% Unable to calculate by menopausal status.	**Comments:** --Single examiner; interobserver variability not an issue --83% followup at 1 year --1 malignancy in patient with normal exam, US, elevated CA-125; menopausal status not reported --Prevalence of abnormal adnexae 1.8% in pre-, 1% in peri-, and 1.4% in postmenopausal women --Normal US in 37.5% of post-, 50% pre- and perimenopausal women --Benign ovarian disease in 20% pre-, 25% peri-, 25% postmenopausal women **Quality assessment:** Reference standard: - (not all got a biopsy) Verification bias: - Test reliability/variability: + Sample size: + Statistical tests: - Blinding: - (not stated) Definition of +/- on screening test: - (subjective)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Jacobs, Stabile, Bridges, et al., 1988 **#6830**	**Geographical location:** London, UK **Dates:** Patients recruited over a 6-month period; published 1988 **Size of population:** 1010 women **Screening study** **Reference standard:** Biopsy or 12-month followup **Reference standard applied to all test negatives?:** All had followup, and a few had biopsy **Test reliability established?:** No **Statistical tests used:** Chi-square **Blinding:** NR **Definition of positive and negative on screening test:** "Palpable pelvic mass of any size that could be clinically distinguished as being separate from the uterus and GI tract"	**Age:** Mean: 54 Range: 45-83 **Menopausal status (n [%]):** Post (> 55): 1010 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** 18 (1.8%) had history of breast cancer **Inclusion criteria:** Age over 45 and postmenopausal **Exclusion criteria:** History of ovarian cancer or undergoing treatment for other cancer; history of bilateral oophorectomy	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	One clinical examiner	1) Abnormal US as gold standard: Dis+ / Dis- / Tot T+: 11 / 17 / 28 T-: 2 / 980 / 982 Tot: 13 / 997 / 1010 Value / Lower 95% CI / Upper 95% CI Se: 84.6% / 65.0% / 100.0% Sp: 98.3% / 97.5% / 99.1% PPV: 39.3% / 21.2% / 57.4% NPV: 99.8% / 99.5% / 100.0% 2) Cancer yes/no: Dis+ / Dis- / Tot T+: 1 / 27 / 28 T-: 0 / 982 / 982 Tot: 1 / 1009 / 1010 Value / Lower 95% CI / Upper 95% CI Se: 100.0% / 0.0% / 100.0% Sp: 97.3% / 96.3% / 98.3% PPV: 3.6% / 0.0% / 10.4% NPV: 100.0% / 99.7% / 100.0%	**Comments:** --"Palpable pelvic mass of any size that could be clinically distinguished as being separate from the uterus and GI tract" **Quality assessment:** Reference standard: + (biopsy and/or 12-month followup) Verification bias: Test reliability/variability: Sample size: - (one cancer) Statistical tests: + Blinding: - (not stated) Definition of +/- on screening test: - (palpable pelvic mass of any size)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Ong, Duffy, and Murphy, 1996 **#780**	**Geographical location:** Dublin, Ireland **Dates:** Jan 1993-Feb 1995 **Size of population:** 86 undergoing laparotomy **Other:** Patients undergoing surgery **Reference standard:** Biopsy **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** NR (retrospective chart review)	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** Pregnant; missing information or no US	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	One gynecologist examiner	1) Ovarian mass yes/no by US: Dis+ / Dis- / Tot T+: **46** / **9** / 55 T-: **18** / **13** / 31 Tot: 64 / 22 / 86 Value (Lower 95% CI, Upper 95% CI) Se: 71.9% (60.9%, 82.9%) Sp: 59.1% (38.5%, 79.6%) PPV: 83.6% (73.9%, 93.4%) NPV: 41.9% (24.6%, 59.3%) 2) Uterine mass: Dis+ / Dis- / Tot T+: **14** / **5** / 19 T-: **4** / **63** / 67 Tot: 18 / 68 / 86 Value (Lower 95% CI, Upper 95% CI) Se: 77.8% (58.6%, 97.0%) Sp: 92.6% (86.4%, 98.9%) PPV: 73.7% (53.9%, 93.5%) NPV: 94.0% (88.4%, 99.7%)	**Comments:** --Separates the Se/Sp for detection of uterine mass and ovarian mass **Quality assessment:** Reference standard: + (biopsy) Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - (not stated) Definition of +/- on screening test: - (not stated)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Padilla, Radosevich, and Milad, 2000 **#460**	**Geographical location:** Chicago, IL **Dates:** Mar 1997-Mar 1998 **Size of population:** 82 adnexal masses in 140 patients undergoing surgery **Other:** Women undergoing laparotomy **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Youden j statistic, Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** Adnexal mass defined as approx. 5 cm or more in greatest diameter	**Age:** Mean: 39.3 **Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR Post (> 55): 14 (10%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Many indications for surgery, including sterilization **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Exam under anesthesia by attendings, residents, and medical students	1) Left adnexa by attending: Dis+ / Dis- / Tot T+: **16** / **9** / 25 T-: **33** / **69** / 102 Tot: 49 / 78 / 127 Se: 32.7% (19.5%–45.8%) Sp: 88.5% (81.4%–95.6%) PPV: 64.0% (45.2%–82.8%) NPV: 67.6% (58.6%–76.7%) 2) Right adnexa by attending: Dis+ / Dis- / Tot T+: **7** / **20** / 27 T-: **26** / **74** / 100 Tot: 33 / 94 / 127 Se: 21.2% (7.3%–35.2%) Sp: 78.7% (70.4%–87.0%) PPV: 25.9% (9.4%–42.5%) NPV: 74.0% (65.4%–82.6%) 3) Left adnexa by resident: Dis+ / Dis- / Tot T+: **18** / **8** / 26 T-: **31** / **81** / 112 Tot: 49 / 89 / 138 Se: 36.7% (23.2%–50.2%) Sp: 91.0% (85.1%–97.0%) PPV: 69.2% (51.5%–87.0%) NPV: 72.3% (64.0%–80.6%) 4) Right adnexa by resident:	**Comments:** --Left and right adnexa were considered separately – abstractor not sure 2x2 tables are correct – don't tell us number of Dis - **Quality assessment:** Reference standard: + (all had surgery) Verification bias: - Test reliability/variability: + Sample size: - (small study) Statistical tests: + Blinding: - (not stated) Definition of +/- on screening test: + (greater than 5 cm adnexa)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	9	14	23
T-	24	91	115
Tot	33	105	138

	Value	Lower 95% CI	Upper 95% CI
Se	27.3%	12.1%	42.5%
Sp	86.7%	80.2%	93.2%
PPV	39.1%	19.2%	59.1%
NPV	79.1%	71.7%	86.6%

5) Left adnexa by student:

	Dis+	Dis-	Tot
T+	11	6	17
T-	38	42	80
Tot	49	48	97

	Value	Lower 95% CI	Upper 95% CI
Se	22.4%	10.8%	34.1%
Sp	87.5%	78.1%	96.9%
PPV	64.7%	42.0%	87.4%
NPV	52.5%	41.6%	63.4%

6) Right adnexa by student:

	Dis+	Dis-	Tot
T+	5	5	10
T-	28	59	87
Tot	33	64	97

	Value	Lower 95% CI	Upper 95% CI
Se	15.2%	2.9%	27.4%
Sp	92.2%	85.6%	98.8%
PPV	50.0%	19.0%	81.0%
NPV	67.8%	58.0%	77.6%

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Padilla, Radosevich, and Milad, 2005 **#7280**	**Geographical location:** Chicago, IL **Dates:** Mar 1997-Mar 1998 **Size of population:** 84 **Screening study** **Registry** **Other** Series of women undergoing gyn surgery **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not discussed **Statistical tests used:** Se, Sp, NPV, PPV, Youden's J statistic, likelihood ratio, logistic regression **Blinding:** Examiners blinded to symptoms, indications **Definition of positive and negative on screening test:** Positive: Adnexal mass ≥ 5 cm	**Age:** Mean (SD): 37.7 (0.93) **Menopausal status (n [%]):** Pre (< 45): 95.2% Peri (45-55): NR Post (> 55): NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Other:** Mean BMI 26.5 18% BMI > 30 **Inclusion criteria:** Women presenting for laparoscopy or laparotomy; range of indications: diagnostic laparoscopy, sterilization, suspected malignancy, etc. **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Examination under anesthesia in dorsal lithotomy position. Bladder drained. Examiners reported adnexal size, presence of mass, uterine position, size, contour, mobility. Examiners divided into board-certified OB/GYN (n = 52), OB/GYN residents (n = 30), 3rd and 4th year med students (n = 40).	1) Total all examiners – detection of adnexal mass: Dis+ / Dis- / Tot T+: 14 / 10 / 24 T-: 76 / 152 / 228 Tot: 90 / 162 / 252 Value (Lower 95% CI – Upper 95% CI) Se: 15.6% (8.1% – 23.0%) Sp: 93.8% (90.1% – 97.5%) PPV: 58.3% (38.6% – 78.1%) NPV: 66.7% (60.5% – 72.8%) (calculated by summing results for atttendings, residents, and students) 2) Attendings – detection of adnexal mass: Dis+ / Dis- / Tot T+: 8 / 4 / 12 T-: 22 / 50 / 72 Tot: 30 / 54 / 84 Value (Lower 95% CI – Upper 95% CI) Se: 28.0% (11.9% – 44.1%) Sp: 93.0% (86.2% – 99.8%) PPV: 66.7% (40.0% – 93.3%) NPV: 69.4% (58.8% – 80.1%) 3) Residents – detection of adnexal mass: Dis+ / Dis- / Tot T+: 5 / 3 / 8 T-: 25 / 51 / 76 Tot: 30 / 54 / 84 Value (Lower 95% CI – Upper 95% CI) Se: 16.0% (2.9% – 29.1%) Sp: 95.0% (89.2% – 100.0%)	**Comments:** --Final diagnoses not presented --Reasons for surgery not systematically presented **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - (reliability not referenced or discussed) Sample size: + (but no a priori sample size presented) Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring

PPV	62.5%	29.0%	96.0%
NPV	67.1%	56.5%	77.7%

4) Students – detection of adnexal mass:

	Dis+	Dis-	Tot
T+	1	3	4
T-	29	51	80
Tot	**30**	**54**	84

	Value	Lower 95% CI	Upper 95% CI
Se	**4.0%**	0.0%	11.0%
Sp	**95.0%**	89.2%	100.0%
PPV	25.0%	0.0%	67.4%
NPV	63.8%	53.2%	74.3%

5) Other:
Likelihood of not detecting an adnexal mass increased with less experience (OR for resident 1.13, student 1.36 compared to attending, although 95% CIs cross 1).

Statistically significant increase in missed diagnosis if subject with BMI > 30 (OR 2.57; 95% CI 1.36 to 4.87), and significant decrease in presence of enlarged uterus (OR 0.48; 95% CI 0.25 to 0.93).

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Roman, Muder-spach, Stein, et al., 1997 **#6160**	**Geographical location:** Los Angeles, CA **Dates:** Jul 1992-Mar 1994 **Size of population:** 226 **Other** Nonconsecutive case series **Reference standard:** Surgical/path findings **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Pearson, logistic regression **Blinding:** No **Definition of positive and negative on screening test:** Positive = fixed, irregular contour, or clinical ascites	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 181 (80.1%) Post (> 55): 45 (19.9%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Suspicious mass needing surgical evaluation **Exclusion criteria:** Emergent laparotomy, clinical or radiologic evidence of metastatic disease, U/S by gynecologist	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	By staff gynecologist in clinic 1 to 4 days prior to surgery	Results not given for 26 women in whom mass was not palpable. 1) All women: Dis+ / Dis- / Tot T+: 22 / 26 / 48 T-: 21 / 131 / 152 Tot: 43 / 157 / 200 Value (Lower 95% CI – Upper 95% CI) Se: 51.2% (36.3% – 66.1%) Sp: 83.6% (77.8% – 89.4%) PPV: 45.8% (31.7% – 59.9%) NPV: 86.2% (80.7% – 91.7%) 2) Premenopausal women: Sufficient data not provided to calculate 2x2 or CIs. Se 50.0% Sp 83.2% PPV 36.1% NPV 89.9% 3) Postmenopausal women: Sufficient data not provided for 2x2 table. Se 53.3% Sp 85.7% PPV 66.7% NPV 77.4%	**Comments:** --Preselected group with "suspicious masses" --Results don't include 26 with nonpalpable masses; data on final diagnosis in these patients not provided --Not clear how low malignant potential tumors were classified in terms of calculation of Se/Sp **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Schutter, Kenemans, Sohn, et al., 1994 **#940**	**Geographical location:** Netherlands and Germany **Dates:** Nov 1990-Dec 1992 **Size of population:** 228 **Other:** Women presenting with pelvic mass **Reference standard:** Biopsy **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Chi-square or Fisher's **Blinding:** NR **Definition of positive and negative on screening test:** Palpable mass of any size, benign vs. malignant	**Age:** Mean: 63 Median: 62 Range: 45-88 **Menopausal status (n [%]):** Post (> 55): 100% **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** > 45 years, postmenopausal **Exclusion criteria:** Another cancer, indeterminate exam	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 199 (87%) **Detected by imaging (n [%]):** 28 (12%) by US **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	One gynecologist examiner	1) Malignant vs. benign: Dis+ / Dis- / Tot T+: **88** / **47** / 135 T-: **7** / **80** / 87 Tot: 95 / 127 / 222 Value (Lower 95% CI – Upper 95% CI) Se: 92.6% (87.4% – 97.9%) Sp: 63.0% (54.6% – 71.4%) PPV: 65.2% (57.1% – 73.2%) NPV: 92.0% (86.2% – 97.7%)	**Quality assessment:** Reference standard: + (all had biopsy) Verification bias: - Test reliability/variability: - Sample size: + (good size) Statistical tests: + Blinding: - (not stated) Definition of +/- on screening test: + (benign or malignant)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
Schutter, Sohn, Kristen, et al., 1998 **#730**	**Geographical location:** Amsterdam, Netherlands; Wurzburg and Mainz, Germany **Dates:** NR (referenced in another paper by this group) **Size of population:** 180 (155 met inclusion/exclusion criteria) **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Lack of data on reliability discussed **Statistical tests used:** Se, Sp, NPV, PPV **Blinding:** NR **Definition of positive and negative on screening test:** Abnormal: mass of any size clinically distinguishable as being separate from uterus and GI tract; examiner asked to	**Age:** Mean: 63 Median: 61 Range: 45-88 **Menopausal status (n [%]):** Post (> 55): 180 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Not described (referenced) **Exclusion criteria:** Not described (referenced)	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Not described; presumably, in outpatient setting	1) Benign vs. malignant (borderline = benign): Dis+ / Dis- / Tot T+: 54 / 24 / 78 T-: 5 / 68 / 73 Tot: 59 / 92 / 151 Value / Lower 95% CI / Upper 95% CI Se: 91.5% / 84.4% / 98.6% Sp: 73.9% / 64.9% / 82.9% PPV: 69.2% / 59.0% / 79.5% NPV: 93.2% / 87.4% / 98.9%	**Comments:** --Examiners not blinded --High prevalence of disease --Clinical history prior to examination not described **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: + (discussed) Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: - (criteria for definition of malignancy not given)

Evidence Table 2 (continued)

Study	Study Design	Patients	Clinical Presentation	Clinical Setting of Exam	Results	Comments/Quality Scoring
	state whether benign or malignant					

Evidence Table 3: *Question 3: Among women with a palpable adnexal mass on exam or a mass identified by ultrasound/imaging, what is the sensitivity/specificity of various evaluation modalities including ultrasound (transvaginal ultrasound, transabdominal ultrasound, color Doppler, 2D vs. 3D ultrasound), CT scan, MRI scan, and CA-125 levels for distinguishing benign from malignant masses?*

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Adonakis, Para-skevaidis, Tsiga, et al., 1996 **#810**	**Geographical location:** Greece **Dates:** Mar 1991- Jun 1993 **Size of population:** 2000/2000 **Screening study** **Reference standard:** Histopathology for selected positives; followup at 12 months for all others **Reference standard applied to all test negatives?:** 180 of the 2000 patients went onto TVUS, and only 35 these had histopathologic diagnosis; 145 verified by clinical followup of repeat exam, TVUS and/or CA-125. No reported loss to followup (although not explicitly stated). **Test reliability established?:** BME – No CA-125 – Yes **Statistical tests used:** Se, Sp **Blinding:** NR, but exams preceded surgery	**Age:** Mean (SD): 58.1(6.9) Range: 45-80 **Menopausal status (n [%]):** Pre: 405 (20%) Peri: 293 (15%) Post: 1302 (65%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women > 45, no evidence of adnexal pathology, agreed to participate **Exclusion criteria:** Women with history of ovarian cancer or any other malignancy, history of bilateral oophorectomy, ascites	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** 50 (3%) + exam 115 (6%) "ambiguous" exam **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Women with elevated CA-125 or and abnormal or ambiguous BME were recalled for TVUS. Only women with +TVUS were referred for further management.	1) CA-125 (T+ ≥ 35 U/ml) Dis+ / Dis- / Tot T+: 3 / 15 / 18 T-: 0 / 1982 / 1982 Tot: 3 / 1997 / 2000 Value (Lower 95% CI, Upper 95% CI) Se: 100.0% (0.0%, 100.0%) Sp: 99.2% (98.9%, 99.6%) PPV: 16.7% (0.0%, 33.9%) NPV: 100.0% (99.8%, 100.0%) 2) PE Dis+ / Dis- / Tot T+: 2 / 57 / 59 T-: 1 / 1940 / 1941 Tot: 3 / 1997 / 2000 Value (Lower 95% CI, Upper 95% CI) Se: 66.7% (13.3%, 100.0%) Sp: 97.1% (96.4%, 97.9%) PPV: 3.4% (0.0%, 8.0%) NPV: 99.9% (99.8%, 100.0%)	**Comments:** --1 tumor LMP grouped in with 2 other malignancies --"Ambiguous" BME was classed as Test -, although all patients with ambiguous BME had TVUS to further evaluate --Borderline tumors considered Dis+ **Quality assessment:** Reference standard: +; all had at least 12 months of followup. Verification bias: +; reference standard of followup applied to all Test reliability/variability: + CA-125, - BME Sample size: -; large sample size, but small number of cases makes CIs around test characteristic estimates wide, especialy for sensitivity Statistical tests: + Blinding: + Definition of +/- on screening test: + CA-125, - BME

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: CA-125 ≥ 35 U/mL BME – "palpable mass"				
Alcazar and Castillo, 2005 **#7460**	**Geographical location:** Pamplona, Spain **Dates:** Jan 2002 – Apr 2004 **Size of population:** 60 patients 69 masses **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp, MCNemar, Fleiss kappa index **Blinding:** Yes **Definition of positive and negative on screening test:** 2D US – presence of at least one of the following:	**Age:** Mean (SD): 48.4 (16.4) Range: 17-82 **Menopausal status (n [%]):** Pre (< 45): 32 (53%) Post (> 55): 28 (47%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women with diagnosis of adnexal mass who received treatment at institution in time frame for who got 2D and 3D US **Exclusion criteria:** "Masses in which the echo features were highly characteristic of a given pathologic condition (such as simple cyst, cystic teratoma, or endometrioma)"	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	All results for masses not patients 1) 2D (combined Doppler and morphology) Dis+ / Dis- / Tot T+: 44 / 3 / 47 T-: 1 / 21 / 22 Tot: 45 / 24 / 69 Value / Lower 95% CI / Upper 95% CI Se: 97.8% / 93.5% / 100.0% Sp: 87.5% / 74.3% / 100.0% PPV: 93.6% / 86.6% / 100.0% NPV: 95.5% / 86.8% / 100.0% 2) 3D (combined) Dis+ / Dis- / Tot T+: 44 / 5 / 49 T-: 1 / 19 / 20 Tot: 45 / 24 / 69 Value / Lower 95% CI / Upper 95% CI Se: 97.8% / 93.5% / 100.0% Sp: 79.2% / 63.0% / 95.4% PPV: 89.8% / 81.3% / 98.3% NPV: 95.0% / 85.4% / 100.0%	**Comments:** --14 of the 60 patients included in this study were included in a previous study by the authors (Alcazar 2003, ref 17) --One person performed all scans (both 2D and 3D); however, he only interpreted the 2D scans. 3D scans interpreted by other individual blinded to 2D results. --Kappa index calculated for interobserver agreement (k = 0.90) --No discussion of clinical pathway to diagnosis. --No discussion of why decision for 3D US – most likely suspicious 2D scan --High incidence of cancer in this study --Descriptive morphologic classification for 2D and doppler– no scoring system used --3D US – definition of positive/negative test not mentioned --Unclear how Doppler included in final table --Unable to stratify by menopausal status --Results reported by masses not patients --LMP tumors grouped in with malignant --Numbers in text and table II don't mesh exactly (2x2 tables here from Table II data) --Authors note no difference in Se

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	thick wall (> 3 mm), thick papillary projections (> 3 mm), solid areas or purely solid echogenicity = complex mass. Doppler – blood flow detected within a papillary projection, solid area, or central area of solid tumor = malignant. 3D – not mentioned.				and Sp between 2D and 3D p = 0.250 --Good discussion of literature on 3D --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: -
Alcazar, Errasti, Zornoza, et al., 1999 **#3110**	**Geographical location:** Spain **Dates:** Jan 1995- Feb 1998 **Size of population:** 94 of 480 **Other** Retrospective case series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Inter and intra assay coefficient for CA-125 reported **Statistical tests used:** Kolmogorov-Smirnov Student t-test Mann-Witney U	**Age:** Mean (SD): 47.4 (16.1) Range: 17-79 **Menopausal status (n [%]):** Pre (< 45): 55.3% Post (> 55): 44.7% **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Sonographically suspicious adnexal mass, "presence of at least one of the following: gross septa (> 3 mm), gross papillary projections (> 3 mm), solid wall nodules, multilocularity, irregular borders or ascitis" Transvaginal color Doppler evaluation and serum CA-125 levels determined prior to surgery	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Color Doppler – all |	Dis+	Dis-	Tot
---	---	---	---
T+	49	6	55
T-	7	32	39
Tot	56	38	94 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	87.5%	78.8%	96.2%
Sp	84.2%	72.6%	95.8%
PPV	89.1%	80.9%	97.3%
NPV	82.1%	70.0%	94.1% | 2) CA-125 – all |	Dis+	Dis-	Tot
---	---	---	---
T+	47	12	59
T-	9	26	35
Tot	56	38	94 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	83.9%	74.3%	93.5%
Sp	68.4%	53.6%	83.2%
PPV	79.7%	69.4%	89.9%
NPV	74.3%	59.8%	88.8%	 3) Doppler – Premenopausal	**Comments:** --LMP tumors grouped in with malignant --Inclusion criteria predispose to increased likelihood of cancer --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: +; pathology Verification bias: +; all underwent surgery Test reliability/variability: +; one reader for Doppler; inter and intra assay coefficient of variation for CA-125 Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Chi-square
McNemars Test
ROC

Definitive histopathological diagnosis

Blinding:
NR

Exclusion criteria:
NR

Definition of positive and negative on screening test:
Doppler
Resistance index [RI=S-D] calculated for each case
Malignancy suspected if lowest RI ≤ 0.45
CA-125 level
CA-125 = 35 U/ml considered as suspicious for malignancy

	Dis+	Dis-	Tot
T+	23	5	28
T-	3	21	24
Tot	26	26	52

	Value	Lower 95% CI	Upper 95% CI
Se	88.5%	76.2%	100.0%
Sp	80.8%	65.7%	95.9%
PPV	82.1%	67.9%	96.3%
NPV	87.5%	74.3%	100.0%

4) CA-125 – Premenopausal

	Dis+	Dis-	Tot
T+	22	8	30
T-	4	18	22
Tot	26	26	52

	Value	Lower 95% CI	Upper 95% CI
Se	84.6%	70.7%	98.5%
Sp	69.2%	51.5%	86.9%
PPV	73.3%	57.5%	89.1%
NPV	81.8%	65.7%	97.9%

5) Doppler – Postmenopausal

	Dis+	Dis-	Tot
T+	26	1	27
T-	4	11	15
Tot	30	12	42

	Value	Lower 95% CI	Upper 95% CI
Se	86.7%	74.5%	98.9%
Sp	91.7%	76.1%	100.0%
PPV	96.3%	89.2%	100.0%
NPV	73.3%	50.9%	95.7%

6) CA-125 – Postmenopausal

	Dis+	Dis-	Tot
T+	25	4	29
T-	5	8	13

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				Tot 30 12 42 Value / Lower 95% CI / Upper 95% CI Se 83.3% 70.0% 96.6% Sp 66.7% 40.0% 93.4% PPV 86.2% 73.6% 98.8% NPV 61.5% 35.0% 88.0%	
Alcazar, Galan, Garcia-Manero, et al., 2003 **#1990**	**Geographical location:** Spain **Dates:** Jun 2001 to Jun 2002 **Size of population:** 44 masses 41 women **Other** Prospective case series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes using kappa **Statistical tests used:** McNemars **Blinding:** Second 3D reviewer blinded, 2D and first 2D reviewer not **Definition of positive and negative on screening test:**	**Age:** Mean (SD): 49.5 Range: 23-75 **Menopausal status (n [%]):** Pre (< 45): 20 (49%) Post (> 55): 21 (51%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women with the diagnosis of complex adnexal masses on 2D TVUS **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) 2D US Dis+ / Dis- / Tot T+ 19 9 28 T- 2 14 16 Tot 21 23 44 Value / Lower 95% CI / Upper 95% CI Se 90.0% 77.2% 100.0% Sp 61.0% 41.1% 80.9% PPV 67.9% 50.6% 85.2% NPV 87.5% 71.3% 100.0% 2) 3D US Dis+ / Dis- / Tot T+ 21 5 26 T- 0 18 18 Tot 21 23 44 Value / Lower 95% CI / Upper 95% CI Se 100.0% 85.7% 100.0% Sp 78.0% 61.1% 94.9% PPV 80.8% 65.6% 95.9% NPV 100.0% 83.3% 100.0%	**Comments:** --Descriptive not numerical/reproducible scoring system --2x2 tables for masses not for individuals --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: +; "state small numbers of patients" Statistical tests: + Blinding: NR Definition of +/- on screening test: +; stated

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Presence of one of the following fulfilled criteria for adnexal mass: A thick wall (> 3 mm) A thick septum (> 3 mm) Thick papillary projections (> 3 mm), solid areas, purely solid echogenicity				
Alcazar and Lopez-Garcia, 2001 **#5740**	**Geographical location:** Spain University Hospital **Dates:** Jun 1998 – May 1999 **Size of population:** 180 women **Other** Consecutive patients undergoing surgery with masses **Reference standard:** Surgery **Reference standard applied to all test negatives?:** No, 180 patients evaluated by TVUS for adnexal mass; only 91 underwent surgery **Test reliability established?:** Yes **Statistical tests used:** ROC Se, Sp McNemar test	**Age:** Mean (SD): 46.6 (14.1) Range: 16-81 **Menopausal status (n [%]):** Pre (< 45): 58 (63.7%) Post (> 55): 33 (36.3%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Adnexal mass undergoing surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Morphology criteria of Sassone et al Dis+ / Dis- / Tot T+: 23 / 19 / 42 T-: 2 / 47 / 49 Tot: 25 / 66 / 91 Value (Lower 95% CI – Upper 95% CI) Se: 92.0% (81.4% – 100.0%) Sp: 71.2% (60.3% – 82.1%) PPV: 54.8% (39.7% – 69.8%) NPV: 95.9% (90.4% – 100.0%) 2) Arterial Doppler (RI) T+≤ 0.45 (in patients in whom arterial flow was detected) (Table 3) Dis+ / Dis- / Tot T+: 17 / 1 / 18 T-: 6 / 26 / 32 Tot: 23 / 27 / 50 Value (Lower 95% CI – Upper 95% CI) Se: 76.0% (58.5% – 93.5%) Sp: 95.5% (87.7% – 100.0%) PPV: 94.4% (83.9% – 100.0%) NPV: 81.3% (67.7% – 94.8%) 3) Venous Doppler; cutoff not described Dis+ / Dis- / Tot T+: 12 / 1 / 13	**Comments:** --LMP tumors considered malignant in analysis --There seems to be an inconsistency between the definition of Venous Doppler; the Se/Sp reported in text and; and the Se/Sp reported in Table 3 --Pre-study history (symptomatic vs asymptomatic) not described --Results for 89 subjects not undergoing surgery not provided **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + for Sasonne and RI Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: + morphology criteria and RI ; - for arterial and venous Doppler

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Blinding: Surgeons blinded to US result Prospective study **Definition of positive and negative on screening test:** Sasonne's scoring system (ref 20) Wall thickness (1-3) Septa (1-3) Inner wall structure (1-4) Echogenicity (1-5) Total score is sum, ranges from 4-15 T+ if score ≥ 9 Arterial flow lowest RI ≤ 0.45 Venous Doppler calculated from ROC curve but no number given in text			T- 6 13 19 Tot 18 14 32 Value / Lower 95% CI / Upper 95% CI Se 68.0% 46.4% 89.6% Sp 93.9% 81.4% 100.0% PPV 92.3% 77.8% 100.0% NPV 68.4% 47.5% 89.3% 4) Venous flow velocity; cutoff 10 cm/s AUC = 0.859 ± 0.06 SEM Dis+ / Dis- / Tot T+ 17 4 21 T- 1 10 11 Tot 18 14 32 Value / Lower 95% CI / Upper 95% CI Se 94.0% 83.0% 100.0% Sp 71.0% 47.2% 94.8% PPV 81.0% 64.2% 97.7% NPV 90.9% 73.9% 100.0%	
Alcazar, Merce, Laparte, et al., 2003 **#5390**	**Geographical location:** Pamplona and Madrid, Spain **Dates:** Part 1 Jan 1995 – Jun 201 Part 2 Jul 2001 – Apr 2002 **Size of population:** Part One 665 (705 masses) Part Two 86 (90 masses) **Other** Part 1 retrospective analysis of ultrasound data to construct scoring system	**Age:** Part 2 Mean (SD): 53.5 (11.3) Range: 20-80 **Menopausal status (n [%]):** Pre (< 45): 26 (30.2%) Post (> 55): 60 (69.8%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Part 1 retrospective analysis of 665 women with adnexal masses who had US in hospital during	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Sassone Dis+ / Dis- / Tot T+ 20 7 27 T- 11 52 63 Tot 31 59 90 Value / Lower 95% CI / Upper 95% CI Se 64.5% 47.7% 81.3% Sp 88.1% 79.8% 96.4% PPV 74.1% 57.5% 90.6% NPV 82.5% 73.2% 91.9% 2) DePriest Dis+ / Dis- / Tot T+ 31 11 42 T- 0 48 48	**Comments:** --Stepwise regression (forward) --Their model not reproducible from description in article --Borderline tumors grouped in with malignant --2x2 tables use cases not individuals --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + for Sassone and DePriest, ? for Ferrazzi, - for current study Sample size: - Statistical tests: +/- Blinding: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Part 2 prospective of consecutive patients to test scoring system and compare with other ultrasound scoring systems **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Sassone – Yes DePriest – Yes Ferrazzi - ? This one - ? **Statistical tests used:** Multivariate logistic regression ROC curves **Blinding:** NR – same individual did all US – prospective part 2 **Definition of positive and negative on screening test:** Not described for Sassone, DePries,t or Ferrazzi (but can assume to be identical to reported in literature). Their scoring system used variables only for: Thick papillary projections, high velocity/low resistance, solid area, and central	time frame. Part 2 – prospective analysis of women (consecutive? – NR) with adnexal masses who had surgery in time frame at hospital **Exclusion criteria:** NR		Tot 31 59 90 Value / Lower 95% CI / Upper 95% CI Se 100.0% 90.3% 100.0% Sp 81.4% 71.5% 91.3% PPV 73.8% 60.5% 87.1% NPV 100.0% 93.8% 100.0% 3) Ferrazzi Dis+ / Dis- / Tot T+ 26 10 36 T- 5 49 54 Tot 31 59 90 Value / Lower 95% CI / Upper 95% CI Se 83.9% 71.0% 96.8% Sp 83.1% 73.5% 92.7% PPV 72.2% 57.6% 86.9% NPV 90.7% 83.0% 98.5% 4) Current study Dis+ / Dis- / Tot T+ 31 3 34 T- 0 56 56 Tot 31 59 90 Value / Lower 95% CI / Upper 95% CI Se 100.0% 90.3% 100.0% Sp 94.9% 89.3% 100.0% PPV 91.2% 81.6% 100.0% NPV 100.0% 94.6% 100.0% 5) Menopause Sassone Dis+ / Dis- / Tot T+ 17 4 21 T- 11 30 41 Tot 28 34 62 Lower / Upper	Definition of +/- on screening test: - for current study

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

flow (unclear how this is measured).

	Value	95% CI	95% CI
Se	61.0%	42.9%	79.1%
Sp	88.0%	77.1%	98.9%
PPV	81.0%	64.2%	97.7%
NPV	73.2%	59.6%	86.7%

6) Menopause DePriest

	Dis+	Dis-	Tot
T+	28	6	34
T-	0	28	28
Tot	28	34	62

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	89.3%	100.0%
Sp	82.4%	69.5%	95.2%
PPV	82.4%	69.5%	95.2%
NPV	100.0%	89.3%	100.0%

7) Menopause Ferrazzi

	Dis+	Dis-	Tot
T+	23	6	29
T-	5	28	33
Tot	28	34	62

	Value	Lower 95% CI	Upper 95% CI
Se	82.0%	67.8%	96.2%
Sp	82.0%	69.1%	94.9%
PPV	79.3%	64.6%	94.1%
NPV	84.8%	72.6%	97.1%

8) Menopause current study

	Dis+	Dis-	Tot
T+	28	2	30
T-	0	32	32
Tot	28	34	62

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	89.3%	100.0%
Sp	94.1%	86.2%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

PPV	93.3%	84.4%	100.0%
NPV	100.0%	90.6%	100.0%

9) Premenopause Sassone

	Dis+	Dis-	Tot
T+	3	3	6
T-	0	22	22
Tot	3	25	28

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	0.0%	100.0%
Sp	88.0%	75.3%	100.0%
PPV	50.0%	10.0%	90.0%
NPV	100.0%	86.4%	100.0%

10) Premenopause DePriest

	Dis+	Dis-	Tot
T+	3	5	8
T-	0	20	20
Tot	3	25	28

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	0.0%	100.0%
Sp	80.0%	64.3%	95.7%
PPV	37.5%	4.0%	71.0%
NPV	100.0%	85.0%	100.0%

11) Premenopause Ferrazzi

	Dis+	Dis-	Tot
T+	3	4	7
T-	0	21	21
Tot	3	25	28

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	0.0%	100.0%
Sp	84.0%	69.6%	98.4%
PPV	42.9%	6.2%	79.5%
NPV	100.0%	85.7%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				12) Premenopause current study Dis+ Dis- Tot T+ 3 1 4 T- 0 24 24 Tot 3 25 28 Value Lower 95% CI Upper 95% CI Se 100.0% 0.0% 100.0% Sp 96.0% 88.3% 100.0% PPV 75.0% 32.6% 100.0% NPV 100.0% 87.5% 100.0%	
Ananda-kumar, Chew, Wong, et al., 1996 **#10980**	**Geographical location:** Singapore **Dates:** 1991-1993 **Size of population:** 146 patients 156 tumors **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Student's t-test Z test of proportions and indices of normality Se, Sp **Blinding:** Not described	**Age:** Mean:For benign 35; for malignant 42.9 Range: for whole study pop 16-71 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women with pelvic tumors detected clinically **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Presence of flow – color imaging alone Dis+ Dis- Tot T+ 26 39 65 T- 8 83 91 Tot 34 122 156 Value Lower 95% CI Upper 95% CI Se 76.5% 62.2% 90.8% Sp 68.0% 59.7% 76.3% PPV 40.0% 28.1% 51.9% NPV 91.2% 85.4% 97.0% 3) RI ≤ 0.66 Dis+ Dis- Tot T+ 24 24 48 T- 10 98 108 Tot 34 122 156 Value Lower 95% CI Upper 95% CI Se 71.0% 55.7% 86.3% Sp 80.3% 73.2% 87.4% PPV 50.0% 35.9% 64.1% NPV 90.7% 85.3% 96.2%	**Comments:** --LMP tumors grouped in with malignant --2x2 tables and results calculated for masses not patients --No description of why 0.66 used as cut point for RI --"Color flow imaging alone" criteria for positive subjective and not reproducible --Overlap in RI range with benign tumors (0.44-0.80) and malignant (0.3-0.60) --From their data, performance of Doppler improves if masses with solid components, however, I was unable to get 2x2 table from these results (Tables 4, 5) --TVUS only **Quality assessment:** Reference standard: +/- Verification bias: + Test reliability/variability: +/- Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: RI <= 0.66 Color flow alone = abnormal vessels "continuously fluctuating rather than pulsatile, also with mosaic pattern with yellow-green color combinations indicating turbulent flow"				
Andolf, Jorgensen, and Astedt, 1990 **#1200**	**Geographical location:** Lund Sweden **Dates:** Oct 1984-Jul 1987 **Size of population:** 801 screened **Screening study** **Reference standard:** Surgery or repeat US or CT within 6 months **Reference standard applied to all test negatives?:** No – but follow up US for all test positives who did not go to surgery **Test reliability established?:** Yes **Statistical tests used:** Not described **Blinding:** Not mentioned	**Age:** Range: 40-70 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history: 190 (23.7%) **Inclusion criteria:** Women 40-70 years old who attended outpatient clinic of OB/GYN university hospital Lund, Sweden **Exclusion criteria:** NR; 6 scans excluded from analysis secondary to poor image quality	**Symptomatic (n [%]):** 419 (52.3%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Using US and BME combined (both positive for test to be positive) |	Dis+	Dis-	Tot
---	---	---	---
T+	6	49	55
T-	0	746	746
Tot	6	795	801 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	50.0%	100.0%
Sp	93.8%	92.2%	95.5%
PPV	10.9%	2.7%	19.1%
NPV	100.0%	99.6%	100.0%		**Comments:** --No description of what constituted an abnormal US --No description of what constituted an abnormal manual exam --Women with normal US and exam – half contacted via mail, cancer cases would have been detected in hospital system, only 2% of them had moved out of cachement area --Six women excluded from results secondary to poor quality scans – no mention of follow up in them (cancer? etc.) --Unable to get 2x2 tables for US and BME by itself --Abdominal US only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: Not described				
Antonic and Rakar, 1995 **#10830**	**Geographical location:** Ljubljana, Slovenia **Dates:** Jan-Jul 1993 **Size of population:** 71 **Prospective case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Fisher exact test Se, Sp Mann-Whitney U test **Blinding:** Not described but prospective **Definition of positive and negative on screening test:** PI [=(S-D)/M] and RI [=(S-D)/S] were calculated but test used presence or absence of colored flow without	**Age:** Premenopausal women: mean 41; range 35-54 Peri 53 (52-53) Post 63 (51-82) **Menopausal status (n [%]):** Pre (< 45): 32 (45.1%) Peri (45-55): 4 (5.6%) Post (> 55): 29(40.8%) 6 had undergone hysterectomy (8.5%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Presence or absence of color flow in mass for all patients Dis+ / Dis- / Tot T+: 16 / 28 / 44 T-: 2 / 25 / 27 Tot: 18 / 53 / 71 Value / Lower 95% CI / Upper 95% CI Se: 88.9% / 74.4% / 100.0% Sp: 47.2% / 33.7% / 60.6% PPV: 36.4% / 22.1% / 50.6% NPV: 92.6% / 82.7% / 100.0% 2) Presence or absence of color flow in mass for menopausal patients Dis+ / Dis- / Tot T+: 13 / 3 / 16 T-: 2 / 11 / 13 Tot: 15 / 14 / 29 Value / Lower 95% CI / Upper 95% CI Se: 86.7% / 69.5% / 100.0% Sp: 78.6% / 57.1% / 100.0% PPV: 81.3% / 62.1% / 100.0% NPV: 84.6% / 65.0% / 100.0% 3) CA-125 for all patients Dis+ / Dis- / Tot T+: 15 / 4 / 19 T-: 3 / 49 / 52 Tot: 18 / 53 / 71 Lower / Upper	**Comments:** --LMP tumors grouped in with malignant --No description of clinical path --Good data on overlap of PI and RI range in malignant and non malignant outcomes. --Did not clearly define visualization of color flow or not --No discussion of inter observer reliability --Combination TVUS and abdominal US used (N for each not specified, unable to stratify) **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: + for CA-125, - for Doppler

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	calculation CA-125 > 34 U/ml				

	Value	95% CI	95% CI
Se	83.3%	66.1%	100.0%
Sp	92.5%	85.3%	99.6%
PPV	78.9%	60.6%	97.3%
NPV	94.2%	87.9%	100.0%

4) CA-125 for postmenopausal patients

	Dis+	Dis-	Tot
T+	13	1	14
T-	2	13	15
Tot	15	14	29

	Value	Lower 95% CI	Upper 95% CI
Se	86.7%	69.5%	100.0%
Sp	92.9%	79.4%	100.0%
PPV	92.9%	79.4%	100.0%
NPV	86.7%	69.5%	100.0%

Asif, Sattar, Dawood, et al., 2004

#1580

Geographical location:
Rawalpindi, Pakistan

Dates: Jan 2001 – Jan 2002

Size of population:
100

Other
Consecutive preoperative patients at hospital with mass

Reference standard:
Histopathology

Reference standard applied to all test negatives?:
Yes

Test reliability established?:
CA-125 yes

Age:
Mean (SD):
For malignant – 45(11)
For B9 – 37(14)

Menopausal status (n [%]):
Pre (< 45): 56(56%)
Peri (45-55):
Post (> 55): 44(44%)

Race/ethnicity (n [%]):
NR

Risk factors (n [%]):
NR

Inclusion criteria:
100 consecutive women admitted to hospital in time frame for surgery for adnexal mass

Exclusion criteria:
NR

Symptomatic (n [%]):
NR

Detected by exam (n [%]):
NR

Detected by imaging (n [%]):
NR

Combination (n [%]):
NR

Additional data used for diagnosis:
NR

1) CA-125 for whole study pop (> 35 U/ml)

	Dis+	Dis-	Tot
T+	45	7	52
T-	10	38	48
Tot	55	45	100

	Value	Lower 95% CI	Upper 95% CI
Se	82.0%	71.8%	92.2%
Sp	84.0%	73.3%	94.7%
PPV	86.5%	77.3%	95.8%
NPV	79.2%	67.7%	90.7%

2) US score (Jacobs) ≥ 1

	Dis+	Dis-	Tot
T+	53	38	91
T-	2	7	9
Tot	55	45	100

	Value	Lower 95% CI	Upper 95% CI
Se	96.4%	91.4%	100.0%

Comments:
--No tests of significance done
--Unable to do 2x2 table for postmenopause (even though have info: 33 cancer, 11 benign – can't assume same test characteristics)
--Pre-study history (symptomatic vs asymptomatic) not described

Quality assessment:
Reference standard: +
Verification bias: +
Test reliability/variability: + (for CA125)
Sample size: -
Statistical tests: +/-
Blinding: -
Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	US - ? **Statistical tests used:** ROC Se, Sp **Blinding:** NR – prospective study **Definition of positive and negative on screening test:** CA-125 – 35U/mL US – Jacobs scoring system: cutoff not described in article			Sp 15.6% 5.0% 26.1% PPV 58.2% 48.1% 68.4% NPV 77.8% 50.6% 100.0% 3) US score (Jacobs) ≥ 2 Dis+ Dis- Tot T+ 42 5 47 T- 13 40 53 Tot 55 45 100 Value Lower 95% CI Upper 95% CI Se 76.4% 65.1% 87.6% Sp 88.9% 79.7% 98.1% PPV 89.4% 80.5% 98.2% NPV 75.5% 63.9% 87.1%	
Balbi, Musone, Menditto, et al., 2001 **#2320**	**Geographical location:** Naples, Italy **Dates:** Jan 1996-Mar 2000 **Size of population:** 92 women **Other** Case series **Reference standard:** Histopathological diagnosis **Reference standard applied to all test negatives?:** No, 18 women with "clearly benign" masses not verified; 2 patients with "clearly malignant" disease (metastases) also excluded. **Test reliability**	**Age:** Range: 40-80 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** "Women with a pelvic mass originating in the ovary" **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Not enough info to stratify by age or menopausal status** 1) PE physical exam by standard protocol. Examiner was asked to predict benign or malignant. Dis+ Dis- Tot T+ 20 13 33 T- 2 37 39 Tot 22 50 72 Value Lower 95% CI Upper 95% CI Se 90.0% 77.5% 100.0% Sp 74.0% 61.8% 86.2% PPV 60.6% 43.9% 77.3% NPV 94.9% 87.9% 100.0% 2) US Dis+ Dis- Tot T+ 19 14 33 T- 3 36 39 Tot 22 50 72	**Comments:** --20 patients excluded for reasons that seem to indicate there wasn't blinding --Physical exam had high sensitivity, but examiners not blinded to patient history or prior diagnosis of pelvic mass --Although RI measured, not included in definition of +US --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: + for CA-125; ? for PE, CA-72-4, and US Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

established?:
Not for PE or CA-72-4
? US and RI
Yes for CA-125

Statistical tests used:
Se, Sp
Multivariate logistic analysis

Blinding:
NR

Definition of positive and negative on screening test:
PE – "palpable mass of any size …clinically distinguishable from the gastrointestinal tract"; clinician asked to designate as benign or malignant
CA-125 > 35 U/ml
CA-72-4 > 3 U/ml
US "multilocular solid tumor or solid tumor" from Valentin et al classification ref 19
RI < 0.4

	Value	Lower 95% CI	Upper 95% CI
Se	**86.0%**	71.5%	100.0%
Sp	**72.0%**	59.6%	84.4%
PPV	57.6%	40.7%	74.4%
NPV	92.3%	83.9%	100.0%

3) CA-125

	Dis+	Dis-	Tot
T+	15	7	22
T-	7	43	50
Tot	**22**	**50**	72

	Value	Lower 95% CI	Upper 95% CI
Se	**68.0%**	48.5%	87.5%
Sp	**86.0%**	76.4%	95.6%
PPV	68.2%	48.7%	87.6%
NPV	86.0%	76.4%	95.6%

4) CA-72-4 > 3 U/ml

	Dis+	Dis-	Tot
T+	13	6	19
T-	9	44	53
Tot	**22**	**50**	72

	Value	Lower 95% CI	Upper 95% CI
Se	**59.0%**	38.4%	79.6%
Sp	**88.0%**	79.0%	97.0%
PPV	68.4%	47.5%	89.3%
NPV	83.0%	72.9%	93.1%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Benjapibal, Sunsaneevitayakul, Boriboonhirunsarn, et al., 2002 **#2150**	**Geographical location:** Bangkok, Thailand **Dates:** Jun 2000-Sep 2001 **Size of population:** 120 7 excluded for no measurable flow **Other:** Consecutive patients **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Student t-test Chi-square analysis **Blinding:** Blinded to ultimate diagnosis, but not to other clinical factors **Definition of positive and negative on screening test:** PI ≤ 1.0 is positive (from Bourne ref 25)	**Age:** Mean (SD): 41 (14) Range: 12-81 **Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR Post (> 55): "one fourth" **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients with suspected ovarian tumors admitted for surgery PI measured **Exclusion criteria:** PI not measurable (7)	**Symptomatic (n [%]):** Abdominal pain 30.8% **Detected by exam (n [%]):** Palpable mass 30% **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) PI ≤ 1.0 is T+ Dis+ / Dis- / Tot T+: 29 / 15 / 44 T-: 6 / 63 / 69 Tot: 35 / 78 / 113 Value / Lower 95% CI / Upper 95% CI Se: 82.9% / 70.4% / 95.4% Sp: 80.8% / 72.1% / 89.5% PPV: 65.9% / 51.9% / 79.9% NPV: 91.3% / 84.7% / 98.0%	**Comments:** --No discussion of 7 excluded (no intent to treat analysis) --One of few studies to describe pre-study clinical history **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Benjapibal, Sunsanee-vitayakul, Boriboon-hirunsarn, et al., 2003 **#5600**	**Geographical location:** Bangkok, Thailand **Dates:** Jul 2001-Jun 2002 **Size of population:** 123 3 excluded **Other** Patients with suspected ovarian tumor admitted for elective surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Chi-square **Blinding:** NR Prospective study (but not blinded to clinical history) **Definition of positive and negative on screening test:** Sonographic score modified from Vera (ref 11) and Kawai (ref 12) positive > 9 (10-14)	**Age:** Mean (SD): 41.5 (14.1) Range: 12-81 **Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR Post (> 55): "one fourth" **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Suspicion of ovarian mass, admitted for surgery **Exclusion criteria:** Non ovarian origin of tumor (n = 3; leiomyoma and parovarian cyst)	**Symptomatic (n [%]):** 92% had gynecological symptoms that made them contact their physicians **Detected by exam (n [%]):** 8% diagnosed at routine gynecological checkup **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Sonographic pattern score ≥ 10 is T+ |	Dis+	Dis-	Tot
---	---	---	---
T+	31	9	40
T-	4	76	80
Tot	35	85	120 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	88.6%	78.1%	99.1%
Sp	89.4%	82.9%	95.9%
PPV	77.5%	64.6%	90.4%
NPV	95.0%	90.2%	99.8%	 Complete data on varying cutoffs provided (ROC curve could be constructed)	**Comments:** --Overlap in dates (3 months) from other study by same group (#2150) --Reliability of scoring system not established --Pre-study clinical history described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Berlanda, Ferrari, Mezzopane, et al., 2002 **#2180**	**Geographical location:** Milan, Italy **Dates:** NR; 3-year period **Size of population:** 215 women 234 masses **Other:** Prospective case series **Reference standard:** Histology **Reference standard applied to all test negatives?:** Assumed yes **Test reliability established?:** NR **Statistical tests used:** Student t-tests Mann Witney U Fischers Exact McNemars **Blinding:** NR **Definition of positive and negative on screening test:** Ferrazzi's morphological score (table 1) Ultrasound 1 = ≤ 3 mm; septa = none, vegetations = none; echogenicity = Sololucent 2 = > 3 mm, septa > 3	**Age:** Median = 60; 47-69 (interquartile) for women with malignant masses Median = 32 and 27-43 for interquartile range **Menopausal status (n [%]):** Pre (< 45): 177 Peri (45-55): Post (> 55): 57 **Race/ethnicity (n [%]):** Italian **Risk factors (n [%]):** NR **Inclusion criteria:** Patients undergoing elective surgical treatment for adnexal masses **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** None	1) Test characteristics based on algorithm Dis+ / Dis- / Tot T+: 27 / 6 / 33 T-: 3 / 198 / 201 Tot: 30 / 204 / 234 Value (95% CI) Se: 90.0% (79.3% - 100.0%) Sp: 97.0% (94.7% - 99.3%) PPV: 81.8% (68.7% - 95.0%) NPV: 98.5% (96.8% - 100.0%) 2) Test characteristics based on morphological score (Ferrazzi) Dis+ / Dis- / Tot T+: 28 / 24 / 52 T-: 2 / 180 / 182 Tot: 30 / 204 / 234 Value (95% CI) Se: 93.0% (83.9% - 100.0%) Sp: 88.0% (83.5% - 92.5%) PPV: 53.8% (40.3% - 67.4%) NPV: 98.9% (97.4% - 100.0%) 3) CA-125 > 35 U/ml Dis+ / Dis- / Tot T+: 22 / 1 / 23 T-: 6 / 23 / 29 Tot: 28 / 24 / 52 Value (95% CI) Se: 78.6% (63.4% - 93.8%) Sp: 95.8% (87.8% - 100.0%) PPV: 95.7% (87.3% - 100.0%) NPV: 79.3% (64.6% - 94.1%) 4) RI ≤ 0.6	**Comments:** --Present number of malignancies for pre and post menopausal but no other information provided to create 2x2. --NOTE: N for 2x2 is masses NOT women. --RI ≤ 0.6 cutoff not explained --Masses with "appearance of cystic teratoma" considered benign regardless of morphology US score --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	mm, low echogenicity 3 = septa > 3 mm, 4 = irregular, mostly solid; vegetations ≤ 3, with echogenic areas 5 = irregular, non-applicable, > 3mm, with heterogeneous echogenic areas, solid Score ≥ 9 considered suspicious for malignancy Additional factors considered Mean diameter ≥ 10 cm; immobility, bilaterality, presence of ascites, resistance index < 0.6 and serum CA-125 > 35 IUml Note: additional factors used to develop an algorithm. Algorithm compared to morphological score. Low risk – masses with score < 9 mm and typical cystic teratomas Moderate risk ≥ 9 suspicious for malignancy, absence of any one of the additional criteria defined above High risk ≥ 9 and any of the above factors.			Dis+ / Dis- / Tot T+: 19 / 6 / 25 T-: 9 / 18 / 27 Tot: 28 / 24 / 52 Value (Lower 95% CI – Upper 95% CI) Se: 67.9% (50.6% – 85.2%) Sp: 75.0% (57.7% – 92.3%) PPV: 76.0% (59.3% – 92.7%) NPV: 66.7% (48.9% – 84.4%)	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Bromley, Goodman, and Benacerraf, 1994 **#4630**	**Geographical location:** Boston, MA **Dates:** Mar 1992-Apr 1993 **Size of population:** 33 **Other** Prospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** NR **Statistical tests used:** None stated Se, Sp, PPV, NPV **Blinding:** NR **Definition of positive and negative on screening test:** Sonography Clear cyst < 3 cm -1 Clear cyst ≥ 3 cm – 2 Cyst with slight irregular wall on one side 3 Cyst with uniform low-level echoes or a single thin septation - 4 Solid ovarian enlargement; cyst with irregular borders,	**Age:** NR **Menopausal status (n [%]):** Post (> 55): 100% **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Pelvic masses diagnosed by sonography and histopathologic verification of disease Consecutive cases **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Scanned transabdominally and transvaginally.	1) Scoring system |	Dis+	Dis-	Tot
---	---	---	---
T+	11	10	21
T-	1	11	12
Tot	12	21	33 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	91.0%	74.8%	100.0%
Sp	52.0%	30.6%	73.4%
PPV	52.4%	31.0%	73.7%
NPV	91.7%	76.0%	100.0% | 2) Resistance index using 0.6 |	Dis+	Dis-	Tot
---	---	---	---
T+	8	4	12
T-	4	17	21
Tot	12	21	33 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	66.0%	39.2%	92.8%
Sp	81.0%	64.2%	97.8%
PPV	66.7%	40.0%	93.3%
NPV	81.0%	64.2%	97.7%		**Comments:** --Unclear where RI 0.6 came from --Borderline tumors in malignant group -- 33 sonographic masses in 1 year seems rather low for tertiary women's hospital --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:- Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	nonspecific ovarian masses 5-6 Multiple septations and nodular cystic mass 7-9 7 = less nodularity; 9 = more nodules and septations Same as 7-9 with ascites 10 Doppler Resistance index = (systolic peak – diastolic trough)/ systolic peak Lowest resistance index used RI < 0.6				
Brown, Doubilet, Miller, et al., 1998 **#3350**	**Geographical location:** Boston, MA **Dates:** Jul 1991-Jul 1996 **Size of population:** 194 **Other** Consecutive patients **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** No (but negatives without surgery had followup US that demonstrated resolution) **Test reliability established?:** Yes	**Age:** Mean (SD): 39.9 (12.7) Range: 16-78 **Menopausal status (n [%]):** Pre (< 45): 135 (69.6%) Post (> 55): 38 (19.6%) don't add to 100% because 21 (10.8%) had hysterectomy **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** All adnexal masses scanned at the institution where both gray-scale and Doppler sonography had been done and **Exclusion criteria:**	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR Malignancy score based on logistic regression model derived from gray-scale and Doppler sonography features Solid component None (0) Hyperechoic (13) non-hyperechoic (394)	1) Malignancy score – cutoff 453 AUC 0.98 ± 0.01 Dis+ / Dis- / Tot T+: 26 / 13 / 39 T-: 2 / 170 / 172 Tot: 28 / 183 / 211 Value / Lower 95% CI / Upper 95% CI Se: 93.0% / 83.5% / 100.0% Sp: 93.0% / 89.3% / 96.7% PPV: 66.7% / 51.9% / 81.5% NPV: 98.8% / 97.2% / 100.0% 2) Malignancy score – cutoff 433 AUC 0.98 ± 0.01 Dis+ / Dis- / Tot T+: 28 / 26 / 54 T-: 0 / 157 / 157 Tot: 28 / 183 / 211 Value / Lower 95% CI / Upper 95% CI	**Comments:** --No model validation --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: - variables in model not specified, no independent validation

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Statistical tests used: Mann-Whitney U test Fisher exact test Chi-square Stepwise logistic regression ROC curves **Blinding:** Yes US done prospectively, scale done after by blinded individual **Definition of positive and negative on screening test:** See column 3	Pregnant masses Premenopausal patients > 10 days after LMP Simple cysts < 2 cm in premenopausal women Extraovarian masses on US	Fluid component (anechoic, echogenic, none) Septations Thin (0) Thick (22) None (38) Wall (thin, thick, none) Free fluid Present (38) Absent (0) Bilateral masses (yes, no) Size, average (cm) Size, maximum (cm) Flow location Central (37) Peripheral only (1) None detected (0)	Se 100.0% 89.3% 100.0% Sp 85.8% 80.7% 90.9% PPV 51.9% 38.5% 65.2% NPV 100.0% 98.1% 100.0%	
Buckshee, Temsu, Bhatla, et al., 1998 **#710**	**Geographical location:** New Delhi, India **Dates:** May 1995-Apr 1997 **Size of population:** 34 individuals 36 tumors **Other** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No for PE Yes for US	**Age:** 20-30: n = 10 31-40: n = 13 41-50: n = 6 > 50: n = 5 **Menopausal status (n [%]):** Pre (< 40): 23 (67.6%) Peri (41-50): 6 (17.6) Post (> 50): 5 (14.7%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history: 1 (2.9%) **Inclusion criteria:** Women with presumed adnexal mass going to surgery **Exclusion criteria:** Women with proven	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) PE (diagnosed as malignant = T+) Dis+ Dis- Tot T+ 7 3 10 T- 2 24 26 Tot 9 27 36 Value Lower 95% CI Upper 95% CI Se 77.8% 50.6% 100.0% Sp 88.9% 77.1% 100.0% PPV 70.0% 41.6% 98.4% NPV 92.3% 82.1% 100.0% 2) TVUS score (≥ 9 indicates T+) Sassone et al. Dis+ Dis- Tot T+ 9 4 13 T- 0 23 23 Tot 9 27 36 Value Lower 95% CI Upper 95% CI	**Comments:** --Unclear how patients were chosen given non-consecutive enrollment --Did blind PE to US --This study validates previous measures PI and Sassone --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Statistical tests used: Fisher's exact test Se, Sp **Blinding:** No but prospective enrollment PE was blinded to US result **Definition of positive and negative on screening test:** PE – clinical impression of benign or malignant Gray-scale sonography – Sassone criteria ≥ 9 malignant PI < 1 malignant	diagnosis of malignancy/metastatic disease on ultrasound or CT		Se 100.0% 66.7% 100.0% Sp 85.2% 71.8% 98.6% PPV 69.2% 44.1% 94.3% NPV 100.0% 87.0% 100.0% 3) Pulsatility Index (< 1 indicates T+) Dis+ Dis- Tot T+ 6 1 7 T- 3 26 29 Tot 9 27 36 Value Lower 95% CI Upper 95% CI Se 66.7% 35.9% 97.5% Sp 96.3% 89.2% 100.0% PPV 85.7% 59.8% 100.0% NPV 89.7% 78.6% 100.0%	
Buist, Golding, Burger, et al., 1994 **#960**	**Geographical location:** The Netherlands **Dates:** Nov 1988-Sep 1992 **Size of population:** 64 **Other** Prospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:**	**Age:** Median: 60 Range: 24-84 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Clinically suspected of having primary or recurrent cancer **Exclusion criteria:** Declined participation, contraindications for one of the diagnostic methods or organizational reasons prevented all methods	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Data are for primary cancer Recurrent presented but not included here. 1) CT – reviewer a Dis+ Dis- Tot T+ 26 10 36 T- 1 8 9 Tot 27 18 45 Value Lower 95% CI Upper 95% CI Se 96.0% 88.6% 100.0% Sp 44.0% 21.1% 66.9% PPV 72.2% 57.6% 86.9% NPV 88.9% 68.4% 100.0% 2) CT – reviewer b Dis+ Dis- Tot T+ 24 3 27 T- 3 15 18 Tot 27 18 45	**Comments:** --PE by gynecological oncologist --Population "clinically suspected of having primary or recurrent ovarian CA"-likely to increase sensitivity of unblinded physical exam --45 with r/o primary cancer, 19 with r/o recurrence (2x2 tables for r/o primary group) --Imaging used descriptive yes/no for cancer – no scoring system used for test --No CA-125 level stated --No PE criteria stated --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: -; wide confidence intervals Statistical tests: + Blinding: + for final results, but

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	ROC curves Fisher's exact Chi-square Se, Sp **Blinding:** Yes **Definition of positive and negative on screening test:** NR	being performed before surgery			examiners all knew patient suspected of having disease Definition of +/- on screening test: -

	Value	Lower 95% CI	Upper 95% CI
Se	**89.0%**	77.2%	100.0%
Sp	**83.0%**	65.6%	100.0%
PPV	88.9%	77.0%	100.0%
NPV	83.3%	66.1%	100.0%

3) MRI – reviewer a

	Dis+	Dis-	Tot
T+	26	12	38
T-	1	6	7
Tot	**27**	**18**	45

	Value	Lower 95% CI	Upper 95% CI
Se	**96.0%**	88.6%	100.0%
Sp	**33.0%**	11.3%	54.7%
PPV	68.4%	53.6%	83.2%
NPV	85.7%	59.8%	100.0%

4) MRI – reviewer b

	Dis+	Dis-	Tot
T+	26	1	27
T-	1	17	18
Tot	**27**	**18**	45

	Value	Lower 95% CI	Upper 95% CI
Se	**96.0%**	88.6%	100.0%
Sp	**94.0%**	83.0%	100.0%
PPV	96.3%	89.2%	100.0%
NPV	94.4%	83.9%	100.0%

5) US

	Dis+	Dis-	Tot
T+	24	10	34
T-	3	8	11
Tot	**27**	**18**	45

	Value	Lower 95% CI	Upper 95% CI
Se	**89.0%**	77.2%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Sp	**44.0%**	21.1%	66.9%
PPV	70.6%	55.3%	85.9%
NPV	72.7%	46.4%	99.0%

6) Physical Exam

	Dis+	Dis-	Tot
T+	26	1	27
T-	1	17	18
Tot	**27**	**18**	45

	Value	Lower 95% CI	Upper 95% CI
Se	**96.0%**	88.6%	100.0%
Sp	**94.0%**	83.0%	100.0%
PPV	96.3%	89.2%	100.0%
NPV	94.4%	83.9%	100.0%

7) CA-125

	Dis+	Dis-	Tot
T+	22	6	28
T-	5	12	17
Tot	**27**	**18**	45

	Value	Lower 95% CI	Upper 95% CI
Se	**81.0%**	66.2%	95.8%
Sp	**67.0%**	45.3%	88.7%
PPV	78.6%	63.4%	93.8%
NPV	70.6%	48.9%	92.2%

Kappa for inter-observer reliability:
CT = 0.28
MRI = 0.41

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Buy, Ghossain, Hugol, et al., 1996 **#4030**	**Geographical location:** Paris, France **Dates:** Jan 1993 – Dec 1994 **Size of population:** 160 patients 115 met inclusion criteria **Other** Prospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Mann Witney McNemar **Blinding:** NR **Definition of positive and negative on screening test:** Sonography Borderline or malignant Echogenic structure against the wall of the cyst present; large irregular homogeneous or heterogeneous echogenic structure Irregular thickened (3 mm) wall or septum	**Age:** Benign Mean: 40 Range: 22-73 Menopausal = 28 Borderline 47 and 50 years Both premenopausal Malignant 57 mean; 22-84 Menopausal =15 **Menopausal status (n [%]):** Pre (< 45): 72 (63%) Post (> 55): 43 (37%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Adnexal mass suspected by physical exam or discovered during previous sonography Only patients who had laparoscopy Laparotomy (not laproscopy) included. **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Conventional sonography – indeterminate masses classified as malignant Dis+ / Dis- / Tot T+: 22 / 16 / 38 T-: 3 / 74 / 77 Tot: 25 / 90 / 115 Value / Lower 95% CI / Upper 95% CI Se: 88.0% / 75.3% / 100.0% Sp: 82.0% / 74.1% / 89.9% PPV: 57.9% / 42.2% / 73.6% NPV: 96.1% / 91.8% / 100.0% 2) Color Doppler and sonography Dis+ / Dis- / Tot T+: 22 / 3 / 25 T-: 3 / 87 / 90 Tot: 25 / 90 / 115 Value / Lower 95% CI / Upper 95% CI Se: 88.0% / 75.3% / 100.0% Sp: 97.0% / 93.5% / 100.0% PPV: 88.0% / 75.3% / 100.0% NPV: 96.7% / 93.0% / 100.0% 3) Resistive Index Dis+ / Dis- / Tot T+: 5 / 3 / 8 T-: 21 / 87 / 108 Tot: 25 / 90 / 116 Value / Lower 95% CI / Upper 95% CI Se: 18.0% / 2.9% / 33.1% Sp: 97.0% / 93.5% / 100.0% PPV: 62.5% / 29.0% / 96.0% NPV: 80.6% / 73.1% / 88.0%	**Comments:** --Unable to stratify --Menopause not defined --No scoring system for US – descriptive only --Borderline on both US and on path grouped in with malignant in analysis --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + **Comments:** --Unable to stratify --Menopause not defined --No scoring system for US – descriptive only --Borderline on both US and on path grouped in with malignant in analysis --Pre-study history (symptomatic vs asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	present. Benign – mass did not present with any of the findings of malignant tumors, or pattern typical of a benign ovarian mass. Method 2. Morphology + color Doppler Presence of color flow in echogenic portion charac. As malignant - considered malignant Absence of color flow – considered benign If mass classified as benign using morphology then malignant if color flow in a regular wall, regular septum or regular solid mass – benign. No color flow – benign. Method 3. Spectral Doppler analysis Absence of arterial flow – considered benign. Measured RI, PI and PSV (no definition provided). Lowest values retained. Mass malignant if Resistive Index ≤ 0.4; Pulsatility Index ≤ 1 and Peak systolic velocity ≥ 15 cm/sec.			4) Pulsatility Index Dis+ Dis- Tot T+ 18 3 21 T- 7 87 94 Tot 25 90 115 Value Lower 95% CI Upper 95% CI Se 71.0% 53.2% 88.8% Sp 97.0% 93.5% 100.0% PPV 85.7% 70.7% 100.0% NPV 92.6% 87.2% 97.9% 5) Peak systolic velocity Dis+ Dis- Tot T+ 12 3 15 T- 13 87 100 Tot 25 90 115 Value Lower 95% CI Upper 95% CI Se 47.0% 27.4% 66.6% Sp 97.0% 93.5% 100.0% PPV 80.0% 59.8% 100.0% NPV 87.0% 80.4% 93.6%	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Canis, Pouly, Wattiez, et al., 1997 **#3710**	**Geographical location:** Clermont-Ferrand, France **Dates:** Jan 1992-Dec 1994 **Size of population:** 558 **Other** Case series **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Se, Sp **Blinding:** No **Definition of positive and negative on screening test:** Ultrasound "suspicious" if solid, mixed, mixed with calcified area, vegetations, cyst wall ≥ 3 mm, thick septa ≥ 3 mm, > 3 septae, multicystic, or ascites"; otherwise considered benign	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** Masses discovered at surgery	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Ultrasound results; low malignant potential = benign Dis+ / Dis- / Tot T+: 29 / 218 / 247 T-: 1 / 310 / 311 Tot: 30 / 528 / 558 Value (Lower 95% CI – Upper 95% CI) Se: 96.7% (90.2% – 100.0%) Sp: 58.7% (54.5% – 62.9%) PPV: 11.7% (7.7% – 15.8%) NPV: 99.7% (99.0% – 100.0%) 2) Low malignant potential = cancer Dis+ / Dis- / Tot T+: 43 / 204 / 247 T-: 2 / 309 / 311 Tot: 45 / 513 / 558 Value (Lower 95% CI – Upper 95% CI) Se: 95.6% (89.5% – 100.0%) Sp: 60.2% (56.0% – 64.5%) PPV: 17.4% (12.7% – 22.1%) NPV: 99.4% (98.5% – 100.0%)	**Comments:** --Clinical history not described --Other test results (CA-125) not given --Not stated whether TVUS or abdominal US **Quality assessment:** Reference standard:+ Verification bias:+ Test reliability/variability: - Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: + **Quality Scoring**

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Carter, Iles, Neven, et al., 1993 **#6370**	**Geographical location:** England **Dates:** NR **Size of population:** 152 **Study type NR** **Reference standard:** Histology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** For CA-125 **Statistical tests used:** None **Blinding:** NR **Definition of positive and negative on screening test:** "95% of normal blood samples have a CA-125 level < 37.2 u/ml" but not necessarily used	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 86 Post (> 55): 66 **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** None **Inclusion criteria:** Presentation with pelvic mass **Exclusion criteria:** NR	**Not specified** **Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA – 125 Not enough data provided to stratify table by menopausal status Dis+ / Dis- / Tot T+: 77 / 1 / 78 T-: 10 / 64 / 74 Tot: 87 / 65 / 152 Value / Lower 95% CI / Upper 95% CI Se: 88.5% / 81.8% / 95.2% Sp: 98.5% / 95.5% / 100.0% PPV: 98.7% / 96.2% / 100.0% NPV: 86.5% / 78.7% / 94.3%	**Comments:** ----Pre-study history (symptomatic vs. asymptomatic) not described --Unclear how patients selected --Recurrent disease included in sample --CA-125 cutoff not clearly defined (35? 37.2?) **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Carter, Lau, Fowler, et al., 1995 **#4240**	**Geographical location:** Minneapolis, MN **Dates:** NR **Size of population:** 123 women **Other** Consecutive **Reference standard:** Histopathology or 12 month followup **Reference standard applied to all test negatives?:** No, but those without operative intervention were followed by US for 12 months **Test reliability established?:** Yes **Statistical tests used:** Se, Sp Chi-squared ROC **Blinding:** NR but prospective study (but not blinded to clinical history) **Definition of positive and negative on screening test:** Calculated from ROC curves Best P I< 1.0 RI < 0.4	**Age:** Mean: 48.3 **Menopausal status (n [%]):** Pre (< 45): 72 (58.5%) Post (> 55): 51(41.5%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women with suspected adnexal mass presenting to University of Minn women's hospital **Exclusion criteria:** NR (everyone else)	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) PI < 1.0 (AUC = 0.732 ± 0.069) Dis+ / Dis- / Tot T+: 13 / 21 / 34 T-: 10 / 79 / 89 Tot: 23 / 100 / 123 Value (95% CI: Lower, Upper) Se: 57.0% (36.8%, 77.2%) Sp: 79.0% (71.0%, 87.0%) PPV: 38.2% (21.9%, 54.6%) NPV: 88.8% (82.2%, 95.3%) 2) RI < 0.4 (AUC = 0.684 ± 0.068) Dis+ / Dis- / Tot T+: 5 / 4 / 9 T-: 18 / 96 / 114 Tot: 23 / 100 / 123 Value (95% CI: Lower, Upper) Se: 22.0% (5.1%, 38.9%) Sp: 96.0% (92.2%, 99.8%) PPV: 55.6% (23.1%, 88.0%) NPV: 84.2% (77.5%, 90.9%) 2x2 table also calculable for PI at cutoffs of 0.6, 0.8 and 1.4 RI at cutoffs of 0.6, 0.8 and 1.0	**Comments:** --No results by menopausal status --Pre-study history (symptomatic vs. asymptomatic) not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: +/- Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Caruso, Caforio, Testa, et al., 1996 **#3810**	**Geographical location:** Rome, Italy **Dates:** NR **Size of population:** 122 **Other** Consecutive patients with diagnosis of adnexal mass scheduled for surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Sasonne's [3] – yes DePriest {1} - ? Valentine [2] - ? "Vascular score" - NO **Statistical tests used:** Student's t test Fisher's exact test **Blinding:** NR but US prior to surgery **Definition of positive and negative on screening test:** Sasonne and DePriest – not described Valentine positive ≥ 3 (where 1 = unilocular cyst, 2 = multilocular	**Age:** Mean (SD): 38.4 (16.5) **Menopausal status (n [%]):** Pre: 88 (70.5%) Post: 36 (29.5%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** 122 consecutive patients with diagnosis of adnexal mass scheduled to undergo surgery at the study hospital **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Sasonne's criteria Dis+ / Dis- / Tot T+: 21 / 27 / 48 T-: 0 / 74 / 74 Tot: 21 / 101 / 122 Value (Lower 95% CI – Upper 95% CI) Se: 100.0% (85.7% – 100.0%) Sp: 73.3% (64.6% – 81.9%) PPV: 43.8% (29.7% – 57.8%) NPV: 100.0% (95.9% – 100.0%) 2) DePriest Score Dis+ / Dis- / Tot T+: 21 / 31 / 52 T-: 0 / 70 / 70 Tot: 21 / 101 / 122 Value (Lower 95% CI – Upper 95% CI) Se: 100.0% (85.7% – 100.0%) Sp: 69.3% (60.3% – 78.3%) PPV: 40.4% (27.0% – 53.7%) NPV: 100.0% (95.7% – 100.0%) 3) Valentin score Dis+ / Dis- / Tot T+: 21 / 39 / 60 T-: 0 / 62 / 62 Tot: 21 / 101 / 122 Value (Lower 95% CI – Upper 95% CI) Se: 100.0% (85.7% – 100.0%) Sp: 61.4% (51.9% – 70.9%) PPV: 35.0% (22.9% – 47.1%) NPV: 100.0% (95.2% – 100.0%) 4) Vascular score	**Comments:** --Aside from Valentin scoring system (which the authors described as "arbitrary"), no description of other scoring systems --Their "vascular scoring system" – mostly subjective measurements save RI of 0.43 as RI cutoff. --Menopause not defined and unable to stratify. Statement in text must be error ("the % of postmenopausal women with benign and malignant lesions was 21 and 71% respectively") --Reported Se/Sp for Sassone criteria (Table 5) do not agree precisely with data reported in Fig 3 **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: ? for Sassonne, De Preist and Valentin (references given but no discussion of reliability) + for "vascular score" – intraobserver CV was calculated for RI portion of the score on 10 patients and was 3.5 (+/-%) Sample size: - Statistical tests: + Blinding: - to clinical history but prospective in that US preceded surgery Definition of +/- on screening test: + for Valentin and vascular Others assumed from literature

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	cyst, 3 = unilocular solid cyst, 4 = multilocular solid tumor, 5 = solid tumor) "Vascular score" – [table1] ≥ 5 is positive (where 1 for vessels present, vascular location 1 for pericystic 2 for in solid part, 2 for randomly dispersed vessels, 2 for "smooth waveform", 2 for lowest RI < 0.430				

	Dis+	Dis-	Tot
T+	21	8	29
T-	0	93	93
Tot	21	101	122

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	85.7%	100.0%
Sp	92.1%	86.8%	97.3%
PPV	72.4%	56.1%	88.7%
NPV	100.0%	96.8%	100.0%

5) Vascular score excluding the 6 patients with a score .+5 studied in luteal phase

	Dis+	Dis-	Tot
T+	18	4	22
T-	0	93	93
Tot	18	97	115

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	83.3%	100.0%
Sp	96.0%	92.1%	99.9%
PPV	81.8%	65.7%	97.9%
NPV	100.0%	96.8%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Chalas, Welshinger, Engellener, et al., 1992 **#5100**	**Geographical location:** Stony Brook, NY **Dates:** May 1980-Apr 1990 **Size of population:** 288 (47 excluded) **Other** Series of patients at single center with pelvic mass who were operated on **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes for CA-125 No for platelet count **Statistical tests used:** Chi-square Se, Sp **Blinding:** NR But prospective **Definition of positive and negative on screening test:** CA-125 > 35U/mL Platelets > 400,000/pico L	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women with pelvic mass diagnosis who underwent surgery in hospital during time frame **Exclusion criteria:** 47 excluded because 1) lack of preop platelet count 2) underlying condition associated with thrombocytosis	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 35 Dis+ / Dis- / Tot T+: 60 / 8 / 68 T-: 14 / 48 / 62 Tot: 74 / 56 / 130 Value (Lower 95% CI – Upper 95% CI) Se: 81.1% (72.2% – 90.0%) Sp: 85.7% (76.5% – 94.9%) PPV: 88.2% (80.6% – 95.9%) NPV: 77.4% (67.0% – 87.8%) 2) Platelets > 400,000/microliter Dis+ / Dis- / Tot T+: 78 / 16 / 94 T-: 61 / 86 / 147 Tot: 139 / 102 / 241 Value (Lower 95% CI – Upper 95% CI) Se: 56.1% (47.9% – 64.4%) Sp: 84.3% (77.3% – 91.4%) PPV: 83.0% (75.4% – 90.6%) NPV: 58.5% (50.5% – 66.5%)	**Comments:** --LMP tumors grouped in with malignant --Although Se, Sp reported for CA-125 and platelets for age > 50, no other numbers reported (no n) ; cannot do 2x2 table; reported Se for CA-125 74% < 50, 85% > 50, Sp 83% < 50, 88% > 50; for thrombocytosis, Se < 50 50%, > 50 60%; Sp < 50 83%, > 50 87% --Se, Sp in abstract differ from those in table VI (which is consistent with calculations from table V) --Clinical presentation not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: +/- Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Chen, Schwartz, and Li, 1990 **#5330**	**Geographical location:** China **Dates:** NR **Size of population:** 188 **Other** Convenience sample **Reference standard:** Histology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Regression analyses Chi-square **Blinding:** NR **Definition of positive and negative on screening test:** CA-125 (serum) > 65 U/ml considered positive	**Age:** 20-42 for healthy blood donors – age not provided for 92 patients with benign pelvic masses and 41 patients with malignant masses of whom 16 had ovarian cancer **Menopausal status** (n [%]): NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** 55 health female blood donors 92 patients with benign pelvic masses 41 patients with malignant masses **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 – serum (among the 92 women with benign masses and 16 patients with ovarian cancer) Dis+ / Dis- / Tot T+: 15 / 37 / 52 T-: 1 / 55 / 56 Tot: 16 / 92 / 108 Value (Lower 95% CI – Upper 95% CI) Se: 93.8% (82.0% – 100.0%) Sp: 59.8% (49.8% – 69.8%) PPV: 28.8% (16.5% – 41.2%) NPV: 98.2% (94.7% – 100.0%)	**Comments:** --Knew in advance what the diagnosis was, so don't know how this impacted outcomes. --CA-125 ≥ 65 = + --Borderline included in malignant --Not prospective --Clinical presentation not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Only 8 ovarian cancers Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Chen, Schwartz, Li, et al., 1988 **#6870**	**Geographical location:** Changsha, China **Dates:** Sep 1985 – Aug 1986 **Size of population:** 211 preoperative 44 normal patients **Other** "Screening" for "normal" patients (but no follow up described) For 211 – diagnosis of mass undergoing surgery at the hospital **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Not to "normal" comparison group but to all of 211 preop patients. **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** NR but testing preceded surgery **Definition of positive and negative on screening test:** CA-125 > 35 (see comments)	**Age:** Mean (SD): Benign masses 38 (11) Malignant 43 (5) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** "Normal" – normal physical exam and LFTs 211 – "pelvic mass" who had surgery **Exclusion criteria:**	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 for 211 operative patients (> 35 U/ml = T+) Dis+ / Dis- / Tot T+: 48 / 61 / 109 T-: 10 / 92 / 102 Tot: 58 / 153 / 211 Value (95% CI) Se: 82.8% (73.1% – 92.5%) Sp: 60.1% (52.3% – 67.9%) PPV: 44.0% (34.7% – 53.4%) NPV: 90.2% (84.4% – 96.0%) 2) CA-125 (> 35 U/ml = T+) limited to patients with epithelial ovarian cancer (excludes non-ovarian malignancies) Dis+ / Dis- / Tot T+: 30 / 61 / 91 T-: 0 / 92 / 92 Tot: 30 / 153 / 183 Value (95% CI) Se: 100.0% (90.0% – 100.0%) Sp: 60.1% (52.3% – 67.9%) PPV: 33.0% (23.3% – 42.6%) NPV: 100.0% (96.7% – 100.0%) 2x2 tables also provided for cutoffs of > 65 U/ml and > 194 U/ml	**Comments:** --No follow up on "normal" patient group (especially the 2 with CA-125 > 35) therefore excluded from 2x2 table --Borderline masses included in malignant group (there were 4) --No description of how subjects chosen (consecutive NR) --Most analyses use CA-125 > 65 as abnormal --This study illustrates the impact of excluding non-ovarian malignancies from the analysis **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - not described Statistical tests: + Blinding: - but testing before surgery Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Chou, Chang, Yao, et al., 1994 **#10930**	**Geographical location:** Taiwan, China **Dates:** Jan 1991 – Feb 1993 **Size of population:** 108 **Case Series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** RI > 0.5 CA-125 > 35U/ml	**Age:** Mean: 38 Range: 11-85 **Menopausal status (n [%]):** Pre (< 45): 89 (82.4%) Post (> 55): 19 (17.6%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Scheduled for surgery in time frame for mass **Exclusion criteria:** 6 excluded : 3 with ovarian CA, 1 with borderline tumor, and 2 with chronic tubal pregnancy	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI < 0.5 |	Dis+	Dis-	Tot
---	---	---	---
T+	22	7	29
T-	3	76	79
Tot	25	83	108 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	88.0%	75.3%	100.0%
Sp	92.0%	86.2%	97.8%
PPV	75.9%	60.3%	91.4%
NPV	96.2%	92.0%	100.0% | 2) CA-125 > 35U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	23	21	44
T-	2	62	64
Tot	25	83	108 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	92.0%	81.4%	100.0%
Sp	75.0%	65.7%	84.3%
PPV	52.3%	37.5%	67.0%
NPV	96.9%	92.6%	100.0% | 3) Combined CA-125 and RI |	Dis+	Dis-	Tot
---	---	---	---
T+	25	2	27
T-	0	81	81
Tot	25	83	108 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	88.0%	100.0%
Sp	97.0%	93.3%	100.0%
PPV	92.6%	82.7%	100.0%
NPV	100.0%	96.3%	100.0%		**Comments:** --5 patients were premenarchal --Unclear why the 6 patients were excluded (did they have a diagnosis of ovarian cancer from previous surgery, etc?) --Clinical pathway not described --Mostly TVUS, however, abdominal used for “those patients who had no sexual experience” – not stated how many there were **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Cohen, Escobar, Scharm, et al., 2001 **#2460**	**Geographical location:** Chicago, IL **Dates:** Apr 1999-Jun 2000 **Size of population:** 71 women **Other** Women referred for surgery at the same hospital **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No – see comments **Statistical tests used:** Se, Sp PPV, NPV **Blinding:** No but US preoperative **Definition of positive and negative on screening test:** "Any multiloculated, complex, or solid mass in which the echo architecture was not highly suggestive of benign histology was categorized as malignant" For 3D Doppler "masses	**Age:** Mean (SD): Pre - 32 Post - 59 Range: 22-80 **Menopausal status (n [%]):** Pre: 40 Post: 31 **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Known "complex" pelvic mass referred for preoperative US **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) 2D TVUS Dis+ / Dis- / Tot T+: 14 / 26 / 40 T-: 0 / 31 / 31 Tot: 14 / 57 / 71 Value (Lower 95% CI – Upper 95% CI) Se: 100.0% (78.6% – 100.0%) Sp: 54.0% (41.1% – 66.9%) PPV: 35.0% (20.2% – 49.8%) NPV: 100.0% (90.3% – 100.0%) 2) 2D plus 3D TVUS Dis+ / Dis- / Tot T+: 14 / 14 / 28 T-: 0 / 43 / 43 Tot: 14 / 57 / 71 Value (Lower 95% CI – Upper 95% CI) Se: 100.0% (78.6% – 100.0%) Sp: 75.0% (63.8% – 86.2%) PPV: 50.0% (31.5% – 68.5%) NPV: 100.0% (93.0% – 100.0%)	**Comments:** --Very poor description of what constituted a positive test – questionable reproducibility --Doppler measurements not done in 2D modality only in 3D – so the study is comparing both 2D to 3D and no Doppler to Doppler. --No use of quantitative Doppler criteria, e.g., RI or PI --No discussion of inter-, intra-observer reliability (especially given poor description of positive test) --1 borderline tumor and 2 metastatic colon cancer included in 14 malignant cases **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: - poor definition

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	with central vascular flow, vascular flow with in excrescences, or flow within septations were graded malignant" based on modified system by Guerriero (ref 9)				
Davies, Jacobs, Woolas, et al., 1993 **#4720**	**Geographical location:** London **Dates:** NR **Size of population:** 124 **Other** Case series - Retrospective review of consecutive analysis of women with diagnosis of mass admitted to hospital for surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** RMI Se, Sp Chi-square Students t test Mann-Whitney U test **Blinding:** NR	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 86 (69.4%) Post (> 55): 38 (30.6%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Consecutive women admitted to hospital for surgery in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 30 U/mL Dis+ / Dis- / Tot T+: 28 / 22 / 50 T-: 9 / 65 / 74 Tot: 37 / 87 / 124 Value / Lower 95% CI / Upper 95% CI Se: 76.0% / 62.2% / 89.8% Sp: 75.0% / 65.9% / 84.1% PPV: 56.0% / 42.2% / 69.8% NPV: 87.8% / 80.4% / 95.3% 2) CA-125 > 50U/mL Dis+ / Dis- / Tot T+: 26 / 13 / 39 T-: 11 / 74 / 85 Tot: 37 / 87 / 124 Value / Lower 95% CI / Upper 95% CI Se: 70.0% / 55.2% / 84.8% Sp: 85.0% / 77.5% / 92.5% PPV: 66.7% / 51.9% / 81.5% NPV: 87.1% / 79.9% / 94.2% 3) US > 1 Dis+ / Dis- / Tot T+: 37 / 62 / 99 T-: 0 / 25 / 25 Tot: 37 / 87 / 124 Lower / Upper	**Comments:** --Standard CA-125 cutoff of 35 not examined --US scoring system for RMI (Jacobs) not often used in other contexts --Clinical presentation not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + ; discussed Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Definition of positive and negative on screening test:
NR – because test of RMI, various cutoff values analyzed

	Value	95% CI	95% CI
Se	**100.0%**	91.9%	100.0%
Sp	**29.0%**	19.5%	38.5%
PPV	37.4%	27.8%	46.9%
NPV	100.0%	88.0%	100.0%

4) US 3

	Dis+	Dis-	Tot
T+	32	23	55
T-	5	64	69
Tot	**37**	**87**	124

	Value	Lower 95% CI	Upper 95% CI
Se	**87.0%**	76.2%	97.8%
Sp	**74.0%**	64.8%	83.2%
PPV	58.2%	45.1%	71.2%
NPV	92.8%	86.6%	98.9%

5) Menopausal status alone

	Dis+	Dis-	Tot
T+	32	36	68
T-	5	51	56
Tot	**37**	**87**	124

	Value	Lower 95% CI	Upper 95% CI
Se	**87.0%**	76.2%	97.8%
Sp	**59.0%**	48.7%	69.3%
PPV	47.1%	35.2%	58.9%
NPV	91.1%	83.6%	98.5%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
DePriest, Gallion, Pavlik, et al., 1997 **#3650**	**Geographical location:** Kentucky **Dates:** Dec 1987-Dec 1993 **Size of population:** 6470 **Screening study** **Reference standard:** Histology and follow up **Reference standard applied to all test negatives?:** Followup applied **Test reliability established?:** NR **Statistical tests used:** Fischers exact test **Blinding:** NR **Definition of positive and negative on screening test:** Ultrasound Ovarian volume >10 cm^3 for postmenopausal women and > 20 cm^3 for premenopausal Cystic tumor with internal papillary or complex projections into its lumen was considered abnormal. Used an algorithm for disease detection	**Age:** Mean: 58 Range: 30-92 **Menopausal status (n [%]):** Numbers of women by menopausal status not specified although used as an entry criterion. **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history: Ovarian cancer = 1597 (24%) Breast cancer = 1976 (30%) Colon cancer = 990 (15%) **Inclusion criteria:** Asymptomatic postmenopausal women > 50 years of age Asymptomatic women > 30 years of age with a documented history of ovarian cancer in at least one primary or secondary relative. **Exclusion criteria:** Known ovarian tumor or personal history of ovarian cancer.	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) TVUS |	Dis+	Dis-	Tot
---	---	---	---
T+	6	84	90
T-	1	6379	6380
Tot	7	6463	6470 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	85.7%	59.8%	100.0%
Sp	98.7%	98.4%	99.0%
PPV	6.7%	1.5%	11.8%
NPV	100.0%	100.0%	100.0% | 2) For operative cases with morphology index < 4 |	Dis+	Dis-	Tot
---	---	---	---
T+	7	34	41
T-	0	49	49
Tot	7	83	90 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	57.1%	100.0%
Sp	59.0%	48.5%	69.6%
PPV	17.1%	5.6%	28.6%
NPV	100.0%	93.9%	100.0%		**Comments:** --Data not provided to stratify by menopausal status --None of the cases with primary ovarian cancer who had CA-125 drawn had level > 35 --DePriest morphology index used 90 operative cases --% followup of normals not described **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: + Statistical tests: + Blinding: +; prospective US Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	If TVUS abnormal, repeat 4-6 weeks. If that's abnormal used additional tests and then surgery. If normal repeat TVUS in one year. If TVUS initially normal, repeat in one year.				
DePriest, Shenson, Fried, et al., 1993 **#6390**	**Geographical location:** Lexington, Kentucky USA University Hospital **Dates:** Jan 1987 – Jan 1992 **Size of population:** 121 **Other** Case series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** T test for means Chi-square and Fisher's exact test for proportions **Blinding:** NR (but prospective) **Definition of positive**	**Age:** Mean (SD): pre 30.9 Post 55.9 Range: pre 3-47 Post 44-74 **Menopausal status (n [%]):** Pre (< 45): 62 (51.2%) Post (> 55): 59 (48.8%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Prospecitive women with mass who had surgery in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) DePriest ≥ 5 Dis+ / Dis- / Tot T+: 13 / 28 / 41 T-: 0 / 80 / 80 Tot: 13 / 108 / 121 Value / Lower 95% CI / Upper 95% CI Se: 100.0% / 76.9% / 100.0% Sp: 74.1% / 65.8% / 82.3% PPV: 31.7% / 17.5% / 46.0% NPV: 100.0% / 96.3% / 100.0%	**Comments:** --Only n for postmenopause reported, unable to do stratified analysis --No discussion of inter/intra observer variability --No discussion of sample size calculation --Clinical presentation not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: +/- Blinding: +/- Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	and negative on screening test: Morphology index score ≥ 5 (Table 1)				
DePriest, van Nagell Jr., Gallion, et al., 1993 **#6880**	**Geographical location:** Kentucky USA University **Dates:** Nov 1987 - June 1992 **Size of population:** 44/3220 **Screening study** **Reference standard:** For women with abnormal TVUS - pathology **Reference standard applied to all test negatives?:** No **Test reliability established?:** No **Statistical tests used:** Fischer's exact test **Blinding:** NR - prospective **Definition of positive and negative on screening test:** US – ovarian volume > 10 cm^3 or "cystic ovarian tumor with a papillary projection into its lumen" Also DePriest score also	**Age:** Mean: 60 Range: 33-90 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history: 502 (15.6%) ovarian CA 1034 (32.1%) breast CA 678 (21.1%) colon CA **Inclusion criteria:** Volunteers for screening program at U of K **Exclusion criteria:** Individuals with prior history of ovarian cancer or pelvic radiation	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** NA **Detected by imaging (n [%]):** NA **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US score ≥ 5 |	Dis+	Dis-	Tot
---	---	---	---
T+	3	6	9
T-	0	15	15
Tot	3	21	24 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	0.0%	100.0%
Sp	71.4%	52.1%	90.8%
PPV	33.3%	2.5%	64.1%
NPV	100.0%	80.0%	100.0%		**Comments:** --Screening study --Test negatives had repeat US in 1 year (don't report compliance with follow up US, or results of those US) --No discussion of reliability of DePriest index **Quality assessment:** Reference standard: + Verification bias: +/- Test reliability/variability: - Sample size: - Statistical tests: +/- Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	used (but unclear how used at what cutoff)				
DePriest, Varner, Powell, et al., 1994 **#10950**	**Geographical location:** USA **Dates:** NR **Size of population:** 213 **Retrospective chart review with re-analysis of US data** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Kappa statistic, Regression analysis **Blinding:** Yes **Definition of positive and negative on screening test:** DePriest morphology index score >=5	**Age:** For benign tumors mean 44.9 with range (16-84) For malignant mean 53.8 (25-78) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** 11 patients excluded due to lack of US or surgical information	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI ≥ 0.5 Dis+ / Dis- / Tot T+: 39 / 46 / 85 T-: 5 / 123 / 128 Tot: 44 / 169 / 213 Value / Lower 95% CI / Upper 95% CI Se: 89.0% / 79.8% / 98.2% Sp: 73.0% / 66.3% / 79.7% PPV: 45.9% / 35.3% / 56.5% NPV: 96.1% / 92.7% / 99.5%	**Comments:** --LMP grouped in with malignant --Good data on reliability/variability --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:+ Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Dowd, Quinn, Rome, et al., 1993 **#4680**	**Geographical location:** Melbourne, AU **Dates:** 1978 to 1989 **Size of population:** 264 patients total although not all had ultrasound, CA-125 and exam results **Other** Retrospective chart review **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** NR **Statistical tests used:** Chi square or Fishers **Blinding:** Tried to predict disease outcome based on clinical exam and ultrasound **Definition of positive and negative on screening test:** CA ≤ 35 u/ml considered normal US impression of reviewer drawn from US report (not film review): "simple, smooth, and/or	**Age:** Range: 15-35 for premenopausal Range: 40 –89 for post **Menopausal status (n [%]):** Pre (< 45): 121 Post (> 55): 143 **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients who had CA-125 performed in presence of pelvic mass **Exclusion criteria:** Test performed for screening purposes only, in absence of pelvic mass excluded Inadequate documentation for pathology.	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 premenopausal Dis+ / Dis- / Tot T+: 32 / 21 / 53 T-: 11 / 57 / 68 Tot: 43 / 78 / 121 Value / Lower 95% CI / Upper 95% CI Se: 74.0% / 60.9% / 87.1% Sp: 73.0% / 63.1% / 82.9% PPV: 60.4% / 47.2% / 73.5% NPV: 83.8% / 75.1% / 92.6% 2) CA-125 post menopausal Dis+ / Dis- / Tot T+: 80 / 9 / 89 T-: 13 / 41 / 54 Tot: 93 / 50 / 143 Value / Lower 95% CI / Upper 95% CI Se: 86.0% / 78.9% / 93.1% Sp: 82.0% / 71.4% / 92.6% PPV: 89.9% / 83.6% / 96.2% NPV: 75.9% / 64.5% / 87.3% 3) CA-125 all patients Dis+ / Dis- / Tot T+: 112 / 30 / 142 T-: 24 / 98 / 122 Tot: 136 / 128 / 264 Value / Lower 95% CI / Upper 95% CI Se: 82.4% / 75.9% / 88.8% Sp: 76.6% / 69.2% / 83.9% PPV: 78.9% / 72.2% / 85.6% NPV: 80.3% / 73.3% / 87.4% 4) Ultrasound all patients	**Comments:** --Unable to construct 2x2 tables for stratified US results; reported values for premenopausal women: Sensitivity 63%, specificity 89%; postmenopausal, sensitivity 87%, specificity 75% --LMP tumors grouped in with malignant --Clinical presentation not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: Sample size: + Statistical tests: + Blinding: - to clinical history Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	unilateral likely to be benign; "solid or mixed consistency, bilateral, irregular or associated ascites…likely malignancy" Clinical exam: "mass hard, irregular, fixed, attached to other structures."			Dis+ / Dis- / Tot T+ 61 / 17 / 78 T- 14 / 90 / 104 Tot 75 / 108 / 183 Value (Lower 95% CI – Upper 95% CI) Se 81.0% (72.1% – 89.9%) Sp 84.0% (77.1% – 90.9%) PPV 78.0% (68.8% – 87.2%) NPV 86.0% (79.3% – 92.7%)	
Einhorn, Bast Jr., Knapp, et al., 1986 **#6860**	**Geographical location:** Sweden **Dates:** Since 1983 – dates unclear **Size of population:** 100 **Other** Retrospective comparison of serum samples with operative outcomes **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** NR	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** Swedish **Risk factors (n [%]):** NR **Inclusion criteria:** Patients with pelvic mass who had surgery For whom banked serum present **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 35 (excluding non-ovarian primary) Dis+ / Dis- / Tot T+ 14 / 9 / 23 T- 4 / 73 / 77 Tot 18 / 82 / 100 Value (Lower 95% CI – Upper 95% CI) Se 77.8% (58.6% – 97.0%) Sp 89.0% (82.3% – 95.8%) PPV 60.9% (40.9% – 80.8%) NPV 94.8% (89.8% – 99.8%) 2) CA-125 > 35 (includes metastatic disease) Dis+ / Dis- / Tot T+ 18 / 5 / 23 T- 5 / 72 / 77 Tot 23 / 77 / 100 Value (Lower 95% CI – Upper 95% CI) Se 78.3% (61.4% – 95.1%) Sp 93.5% (88.0% – 99.0%) PPV 78.3% (61.4% – 95.1%) NPV 93.5% (88.0% – 99.0%) 2) CA-125 > 35 (classifying borderline as "benign"	**Comments:** --Borderline tumors included in malignant --Slight difference in 2x2 table specificity calculated here (89%) and from text (93%) --No statistical tests of significance **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: CA-125 > 35 U/ml			Dis+ Dis- Tot T+ 16 7 23 T- 5 72 77 Tot 21 79 100 Value / Lower 95% CI / Upper 95% CI Se 76.2% 58.0% 94.4% Sp 91.1% 84.9% 97.4% PPV 69.6% 50.8% 88.4% NPV 93.5% 88.0% 99.0%	
Ekerhovd, Wienerroith, Staudach, et al, 2001 **#8780**	**Geographical location:** Salzburg, Austria, and Goteborg, Sweden **Dates:** Jan 1992-Dec 1997 **Size of population:** 1304 **Other** Case series of all women with unilocular adnexal cyst on transvaginal US **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** t-test, chi-square **Blinding:**	**Age:** Range: 14-90 NR for entire group **Menopausal status (n [%]):** Pre (< 45): 927 (71.1%) Post (> 55):377 (28.9%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Scheduled for surgery and unilocular cyst **Exclusion criteria:** Presence of internal septae	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Low malignant potential = benign, presence of solid areas or papillations = positive test Dis+ Dis- Tot T+ 13 631 644 T- 4 656 660 Tot 17 1287 1304 Value / Lower 95% CI / Upper 95% CI Se 76.5% 56.3% 96.6% Sp 51.0% 48.2% 53.7% PPV 2.0% 0.9% 3.1% NPV 99.4% 98.8% 100.0% 2) Low malignant potential = cancer Dis+ Dis- Tot T+ 24 620 644 T- 7 653 660 Tot 31 1273 1304 Value / Lower 95% CI / Upper 95% CI Se 77.4% 62.7% 92.1% Sp 51.3% 48.6% 54.0% PPV 3.7% 2.3% 5.2% NPV 98.9% 98.2% 99.7%	**Comments:** --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: + Statistical tests:+ Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: Suspicious: unilocular with small solid areas or papillary formation Benign: Simple cysts				
Fenchel, Grab, Nuessle, et al., 2002 **#2220**	**Geographical location:** Ulm, Germany University hospital **Dates:** May 1997 – Feb 1999 **Size of population:** 99 women **Other** Consecutive patients referred to hospital **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** 97 had histopathology, 1 had cytology 1 US follow up **Test reliability established?:** No for PET Yes for US ? for MRI **Statistical tests used:** Se, Sp, PPV, NPV **Blinding:** Yes	**Age:** Mean (SD):46(15) Range: 18-83 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Consecutive, asymptomatic "sonographically suspect" (by referring physician adnexal mass referred to hospital **Exclusion criteria:** Pregnant women, clinical symptoms of malignancy, under 18 years	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Combined US and Doppler |	Dis+	Dis-	Tot
---	---	---	---
T+	11	35	46
T-	1	52	53
Tot	12	87	99 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	92.0%	76.7%	100.0%
Sp	60.0%	49.7%	70.3%
PPV	23.9%	11.6%	36.2%
NPV	98.1%	94.5%	100.0% | 2) MRI |	Dis+	Dis-	Tot
---	---	---	---
T+	10	15	25
T-	2	72	74
Tot	12	87	99 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	83.0%	61.7%	100.0%
Sp	83.0%	75.1%	90.9%
PPV	40.0%	20.8%	59.2%
NPV	97.3%	93.6%	100.0% | 3) FDG PET |	Dis+	Dis-	Tot
---	---	---	---
T+	7	21	28
T-	5	66	71
Tot	12	87	99		**Comments:** --Three different US scores used (DePriest, Kawai,and RI) – although each is well described, how each contributed to the overall diagnosis for this study is not discussed (used in series, or in parallel?) --Hospital referrals – not population-based --Borderline tumors (LMP = 2) probably included in malignant category (unclear – but no examples of borderline in benign tumor descriptions) --**May be same patient population as Grab #2720** **Quality assessment:** Reference standard: +/-; length of time for followup for one non-surgical case not described Verification bias: -; not discussed Test reliability/variability: + for component US tests, however it is unclear how these were grouped together for this study's single diagnostic assessment Other tests - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: For PET – "interpreted visually in consensus" by "2 or 3 experiences nuclear med physicians For FDG uptake – "subjective" scale US DePriest,(≥ 5), Kawai (9-12 = malignant) and Doppler RI < 0.45 = malignant			Se 58.0% (30.1% – 85.9%) Sp 76.0% (67.0% – 85.0%) PPV 25.0% (9.0% – 41.0%) NPV 93.0% (87.0% – 98.9%)	
Ferdeghini, Gadducci, Prontera, et al., 1993 **#4710**	**Geographical location:** Italy **Dates:** NR **Size of population:** 183 **Other** 2 retrospective samples: one if cancer one if benign – both consecutive **Reference standard:** Histology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Student t test Chi square Fishers exact test	**Age:** Median (with range): Ovarian cancer = 60 (35-91) Benign = 35 (13 – 76) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Had pre-operative levels of SIL-2R and CA-125 **Exclusion criteria:** Autoimmune or rheumatic disease	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 Dis+ / Dis- / Tot T+: 42 / 9 / 51 T-: 12 / 120 / 132 Tot: 54 / 129 / 183 Value (Lower 95% CI – Upper 95% CI) Se 77.8% (66.7% – 88.9%) Sp 93.0% (88.6% – 97.4%) PPV 82.4% (71.9% – 92.8%) NPV 90.9% (86.0% – 95.8%)	**Comments:** --CA-125 ≥ 83 U/ml **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: + Statistical tests:+ Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Blinding: NR **Definition of positive and negative on screening test:** CA-125 ≥ 83 U/ml.				
Ferrazzi, Zanetta, Dordoni, et al., 1997 **#3570**	**Geographical location:** Milan, Italy University **Dates:** 1995-96 (2 yrs) **Size of population:** 330 masses **Other** Case series in multi-center **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Sassone – yes DePriest – yes This study – no **Statistical tests used:** ROC curve **Blinding:** NR - prospective **Definition of positive and negative on screening test:** Sassone (per original	**Age:** Mean (SD): 45 (16) Range: 19-89 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Surgery within 7 days of US, detailed pathology available, women with mass in time frame at three hospitals in Italy **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) This study > 9 Dis+ / Dis- / Tot T+: 60 / 86 / 146 T-: 9 / 175 / 184 Tot: 69 / 261 / 330 Value / Lower 95% CI / Upper 95% CI Se: 87.0% / 79.1% / 94.9% Sp: 67.0% / 61.3% / 72.7% PPV: 41.1% / 33.1% / 49.1% NPV: 95.1% / 92.0% / 98.2% 2) Sassone > 9 Dis+ / Dis- / Tot T+: 51 / 91 / 142 T-: 18 / 170 / 188 Tot: 69 / 261 / 330 Value / Lower 95% CI / Upper 95% CI Se: 74.0% / 63.7% / 84.3% Sp: 65.0% / 59.2% / 70.8% PPV: 35.9% / 28.0% / 43.8% NPV: 90.4% / 86.2% / 94.6% 3) DePriest Dis+ / Dis- / Tot T+: 61 / 157 / 218 T-: 8 / 104 / 112 Tot: 69 / 261 / 330 Lower / Upper	**Comments:** --No discussion of inter/intra observer reliability variability with this new scoring system --No power calculation for study --Good use of ROC curves and testing between curves **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results				Comments/Quality Scoring
	article) > 9 DePriest (per original article) > 5 This study – Table 2 > 9				Value	95% CI	95% CI	
				Se	**88.0%**	80.3%	95.7%	
				Sp	**40.0%**	34.1%	45.9%	
				PPV	28.0%	22.0%	33.9%	
				NPV	92.9%	88.1%	97.6%	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Finkler, Benacerraf, Lavin, et al., 1988 **#1230**	**Geographical location:** Boston, MA **Dates:** Nov 1986 to Apr 1987 **Size of population:** 131 consecutive patients 106 eventually retained **Other** Prospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** NR **Statistical tests used:** Fisher's exact **Blinding:** Yes **Definition of positive and negative on screening test:** CA-125 > 35 U/mL considered positive US had two evaluations first (Table 1) Finkler score ≥ 7 = malignant second Primary US = impression only	**Age:** Mean: 45.2 Range: 17-84 **Menopausal status (n [%]):** 74 Pre (< 45): Peri (45-55): Post (> 55): **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Ovarian mass who were scheduled to under exploratory laparotomy Had a pre-operative ultrasound Consecutive patients **Exclusion criteria:** Original ultrasound unavailable or uninterpretable. Pregnant or with histologic cancer diagnosis	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Original ultrasound – premenopausal Dis+ Dis- Tot T+ 2 2 4 T- 16 54 70 Tot 18 56 74 Value / Lower 95% CI / Upper 95% CI Se 11.0% 0.0% 25.5% Sp 96.0% 90.9% 100.0% PPV 50.0% 1.0% 99.0% NPV 77.1% 67.3% 87.0% 2) Specialist ultrasound – premenopausal T+ 9 2 11 T- 9 54 63 Tot 18 56 74 Value / Lower 95% CI / Upper 95% CI Se 50.0% 26.9% 73.1% Sp 96.0% 90.9% 100.0% PPV 81.8% 59.0% 100.0% NPV 85.7% 77.1% 94.4% 3) CA-125 – premenopausal Dis+ Dis- Tot T+ 9 17 26 T- 9 39 48 Tot 18 56 74 Value / Lower 95% CI / Upper 95% CI Se 50.0% 26.9% 73.1% Sp 69.0% 56.9% 81.1% PPV 34.6% 16.3% 52.9% NPV 81.3% 70.2% 92.3% 4) Original US – postmenopausal	**Comments:** --Original US based on impression of cancer vs. benign only --"Specialist" US used scoring system --Unclear if "specialist" US was blinded --Abdominal US – no TVUS --CA-125 significantly improved positive and negative predictive values in postmenopausal women when added to clinical impression or prior ultrasound **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - ; underpowered Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	9	2	11
T-	10	11	21
Tot	19	13	32

	Value	Lower 95% CI	Upper 95% CI
Se	47.0%	24.6%	69.4%
Sp	85.0%	65.6%	100.0%
PPV	81.8%	59.0%	100.0%
NPV	52.4%	31.0%	73.7%

5) Specialist US – postmenopausal

	Dis+	Dis-	Tot
T+	15	1	16
T-	4	12	16
Tot	19	13	32

	Value	Lower 95% CI	Upper 95% CI
Se	78.0%	59.4%	96.6%
Sp	92.0%	77.3%	100.0%
PPV	93.8%	81.9%	100.0%
NPV	75.0%	53.8%	96.2%

6) CA-125 – postmenopausal

	Dis+	Dis-	Tot
T+	16	1	17
T-	3	12	15
Tot	19	13	32

	Value	Lower 95% CI	Upper 95% CI
Se	84.0%	67.5%	100.0%
Sp	92.0%	77.3%	100.0%
PPV	94.1%	82.9%	100.0%
NPV	80.0%	59.8%	100.0%

7) CA-125 – all ages combined

	Dis+	Dis-	Tot
T+	25	18	43
T-	12	51	63

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Tot 37 69 106

	Value	Lower 95% CI	Upper 95% CI
Se	67.6%	52.5%	82.7%
Sp	73.9%	63.6%	84.3%
PPV	58.1%	43.4%	72.9%
NPV	81.0%	71.3%	90.6%

8) US total

	Dis+	Dis-	Tot
T+	24	55	79
T-	22	68	90
Tot	46	123	169

	Value	Lower 95% CI	Upper 95% CI
Se	52.2%	37.7%	66.6%
Sp	55.3%	46.5%	64.1%
PPV	30.4%	20.2%	40.5%
NPV	75.6%	66.7%	84.4%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Fleischer, Rodgers, Kepple, et al., 1992 **#6460**	**Geographical location:** Nashville, TN **Dates:** NR **Size of population:** 62 **Other** Case series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** PI < 1.0	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR – mass – surgery - US **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Doppler Dis+ Dis- Tot T+ 17 1 18 T- 3 41 44 Tot 20 42 62 Value / Lower 95% CI / Upper 95% CI Se 85.0% 69.4% 100.0% Sp 97.6% 93.0% 100.0% PPV 94.4% 83.9% 100.0% NPV 93.2% 85.7% 100.0%	**Comments:** --2x2 tables different if pull data from text or from Table 2 --Table 2 and text confuse positive/negative predictive value and sensitivity/specificity **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: +/- Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Franchi, Beretta, Ghezzi, et al., 1995 **#6270**	**Geographical location:** Italy **Dates:** Jan 1991 to Dec 1993 **Size of population:** 129 **Screening study** Case series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** NR **Statistical tests used:** Mann-Witney U ROC curves **Blinding:** US blinded to lab results **Definition of positive and negative on screening test:** 2D Ultrasound: Maximum diameter (5 cm), solid areas, high echogenicity, multilocular appearance, irregular borders, papillary intracystic vegetations, presence of ascites Color Doppler Imaging RI = systolic peak –	**Age:** Median: 44 Range: 12-91 **Menopausal status (n [%]):** Pre (< 45): 83 (64.3%) Peri (45-55): NR Post (> 55): NR **Race/ethnicity (n [%]):** Italian **Risk factors (n [%]):** NR **Inclusion criteria:** Abnormal findings on pelvic exam and 2D sonographic features of adnexal mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Premenopausal - CA-125 Dis+ / Dis- / Tot T+: 8 / 26 / 34 T-: 3 / 46 / 49 Tot: 11 / 72 / 83 Value (Lower 95% CI – Upper 95% CI) Se: 72.7% (46.4% – 99.0%) Sp: 63.8% (52.7% – 74.9%) PPV: 23.5% (9.3% – 37.8%) NPV: 93.9% (87.2% – 100.0%) 2) Premenopausal - Sonography Dis+ / Dis- / Tot T+: 8 / 10 / 18 T-: 3 / 62 / 65 Tot: 11 / 72 / 83 Value (Lower 95% CI – Upper 95% CI) Se: 72.7% (46.4% – 99.0%) Sp: 86.1% (78.1% – 94.1%) PPV: 44.4% (21.5% – 67.4%) NPV: 95.4% (90.3% – 100.0%) 3) Combined ages Dis+ / Dis- / Tot T+: 31 / 15 / 46 T-: 6 / 77 / 83 Tot: 37 / 92 / 129 Value (Lower 95% CI – Upper 95% CI) Se: 83.8% (71.9% – 95.7%) Sp: 83.7% (76.1% – 91.2%) PPV: 67.4% (53.8% – 80.9%) NPV: 92.8% (87.2% – 98.3%) 4) Premenopausal - Color Doppler Imaging	**Comments:** --ROC curves used to generate RI cutoff --CA-125 > 40 U/ml --No US scoring system – descriptive only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

diastolic peak/systolic peak. Artery with lowest RI considered to indicate the malignant potential of the tumor.
RI < 0.65

CA-125 ≥ 40UI/ml suspected of malignant pelvic tumor

	Dis+	Dis-	Tot
T+	9	20	29
T-	2	52	54
Tot	11	72	83

	Value	Lower 95% CI	Upper 95% CI
Se	81.8%	59.0%	100.0%
Sp	72.2%	61.9%	82.5%
PPV	31.0%	14.2%	47.9%
NPV	96.3%	91.3%	100.0%

5) Color Doppler Imaging - combined ages

	Dis+	Dis-	Tot
T+	30	26	56
T-	7	66	73
Tot	37	92	129

	Value	Lower 95% CI	Upper 95% CI
Se	81.1%	68.5%	93.7%
Sp	71.7%	62.5%	80.9%
PPV	53.6%	40.5%	66.6%
NPV	90.4%	83.7%	97.2%

6) Postmenopausal – CA-125

	Dis+	Dis-	Tot
T+	20	3	23
T-	6	17	23
Tot	26	20	46

	Value	Lower 95% CI	Upper 95% CI
Se	76.9%	60.7%	93.1%
Sp	85.0%	69.4%	100.0%
PPV	87.0%	73.2%	100.0%
NPV	73.9%	56.0%	91.9%

7) All ages CA-125

	Dis+	Dis-	Tot
T+	28	29	57
T-	9	63	72

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Tot	37	92	129

	Value	Lower 95% CI	Upper 95% CI
Se	75.7%	61.9%	89.5%
Sp	68.5%	59.0%	78.0%
PPV	49.1%	36.1%	62.1%
NPV	87.5%	79.9%	95.1%

8) Postmenopausal - Sonography

	Dis+	Dis-	Tot
T+	23	5	28
T-	3	15	18
Tot	26	20	46

	Value	Lower 95% CI	Upper 95% CI
Se	88.5%	76.2%	100.0%
Sp	75.0%	56.0%	94.0%
PPV	82.1%	68.0%	96.3%
NPV	83.3%	66.1%	100.0%

9) Postmenopausal - Color Doppler Imaging

	Dis+	Dis-	Tot
T+	21	6	27
T-	5	14	19
Tot	26	20	46

	Value	Lower 95% CI	Upper 95% CI
Se	81.8%	67.0%	96.6%
Sp	72.2%	52.6%	91.8%
PPV	77.8%	62.1%	93.5%
NPV	73.7%	53.9%	93.5%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Gadducci, Baicchi, Marrai, et al., 1996 **#6230**	**Geographical location:** Pisa, Italy University Hospital **Dates:** NR **Size of population:** 124 women (3 excluded = 121) **Other** Consecutive patients referred for surgery with diagnosis of ovarian mass **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Mann-Whitney U test Spearman rank correlation test Logistic regression ROC curves **Blinding:** No – but consecutive enrollment **Definition of positive and negative on screening test:** CA-125 > 65 U/ml D-Dimer > 416 ng/ml	**Age:** NR **Menopausal status (n [%]):** Pre: 57 (47.1%) Post: 64 (52.9%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Consecutive women with clinical diagnosis of ovarian mass to undergo surgery **Exclusion criteria:** Cardiovascular disease, diabetes, acute or chronic inflammatory disease, previous malignancy, or previous episodes of thrombophlebitis or thromboembolia. 2 patients excluded for uterine fibroid (after surgery) 1 excluded for leiomyosarcoma, of small bowel	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 65 U/ml – premenopause |	Dis+	Dis-	Tot
---	---	---	---
T+	8	4	12
T-	4	41	45
Tot	12	45	57 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	66.7%	40.0%	93.4%
Sp	91.1%	82.8%	99.4%
PPV	66.7%	40.0%	93.3%
NPV	91.1%	82.8%	99.4% | 2) D-Dimer - premenopause |	Dis+	Dis-	Tot
---	---	---	---
T+	12	4	16
T-	0	41	41
Tot	12	45	57 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	75.0%	100.0%
Sp	91.1%	82.8%	99.4%
PPV	75.0%	53.8%	96.2%
NPV	100.0%	92.7%	100.0% | 3) D-Dimer – combined ages |	Dis+	Dis-	Tot
---	---	---	---
T+	51	11	62
T-	5	54	59
Tot	56	65	121 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	91.1%	83.6%	98.5%
Sp	83.1%	74.0%	92.2%
PPV	82.3%	72.7%	91.8%
NPV	91.5%	84.4%	98.6%	 4) CA-125 > 65 U/ml - postmenopause	**Comments:** **--Most of the 124 patients in this study were included in Gadducci et al., 1988 (#6650)** --Age breakdown or definition of menopause not described – however, this article stratifies results by menopausal status. --Cutoff for CA-125 is > 65 U/ml --D-Dimer - cutoff had been previously evaluated in other study (using most of the same patients) by same authors [Reference 22] --D-dimer performance characteristic likely overestimated since these data are not independent of the data used to select cutoff value. **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - not discussed Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	35	0	35
T-	9	20	29
Tot	44	20	64

	Value	Lower 95% CI	Upper 95% CI
Se	79.5%	67.6%	91.4%
Sp	100.0%	85.0%	100.0%
PPV	100.0%	91.4%	100.0%
NPV	69.0%	52.1%	85.8%

5) CA-125 > 65 U/ml – all ages

	Dis+	Dis-	Tot
T+	43	4	47
T-	13	61	74
Tot	56	65	121

	Value	Lower 95% CI	Upper 95% CI
Se	76.8%	65.7%	87.8%
Sp	93.8%	88.0%	99.7%
PPV	91.5%	83.5%	99.5%
NPV	82.4%	73.8%	91.1%

6) D-Dimer - postmenopause

	Dis+	Dis-	Tot
T+	39	7	46
T-	5	13	18
Tot	44	20	64

	Value	Lower 95% CI	Upper 95% CI
Se	88.6%	79.2%	98.0%
Sp	65.0%	44.1%	85.9%
PPV	84.8%	74.4%	95.2%
NPV	72.2%	51.5%	92.9%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Gadducci, Capriello, Bartolini, et al., 1988 **#6650**	**Geographical location:** Pisa Italy University Hospital **Dates:** NR **Size of population:** 119 women **Other** Patients undergoing surgery for mass **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** For CA-125 – yes For US – no **Statistical tests used:** Sensitivity, specificity **Blinding:** No – probably prospective **Definition of positive and negative on screening test:** US scoring system ≥ 10 (of 16 with 4 points for shape, ascites, outline, and structure)	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients undergoing surgery for mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 ≥ 35 U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	31	20	51
T-	5	63	68
Tot	36	83	119 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	86.1%	74.8%	97.4%
Sp	75.9%	66.7%	85.1%
PPV	60.8%	47.4%	74.2%
NPV	92.6%	86.4%	98.9% | 2) US |	Dis+	Dis-	Tot
---	---	---	---
T+	26	2	28
T-	10	81	91
Tot	36	83	119 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	72.2%	57.6%	86.8%
Sp	97.9%	94.8%	100.0%
PPV	92.9%	83.3%	100.0%
NPV	89.0%	82.6%	95.4%		**Comments:** --US scoring system described but not grounded – appears to be a unique (hospital specific? operator specific?) scoring system – also unclear how cutoff of ≥ 10 was fixed --CA-125 cutoff ≥ 65 U/ml preferred by authors, but 2x2 table reported only for ≥ 35 U/ml as that is what is in common clinical practice. **--Patient data overlaps with article Gadducci et al., 1996 (#6230)** --Referral criteria etc. not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - especially given the novel US scoring system Sample size: - Statistical tests: + Blinding: ? Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Gadducci, Ferdeghini, Prontera, et al., 1992 **#6850**	**Geographical location:** Italy **Dates:** NR **Size of population:** 344 **Other** Consecutive case series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes for CA-125 **Statistical tests used:** Chi-square **Blinding:** NR **Definition of positive and negative on screening test:** CA-125 35 and 65 U/ml	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** Italian **Risk factors (n [%]):** NR **Inclusion criteria:** Patients undergoing laparotomy for ovarian masses **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 ≥ 65 U/ml (Age < 50 years) Dis+ / Dis- / Tot T+: 8 / 146 / 154 T-: 8 / 51 / 59 Tot: 16 / 197 / 213 Value (95% CI): Se 50.0% (25.5%–74.5%); Sp 26.0% (19.9%–32.1%); PPV 5.2% (1.7%–8.7%); NPV 86.4% (77.7%–95.2%) 2) CA-125 ≥ 65U/ml (Age ≥ 50) Dis+ / Dis- / Tot T+: 60 / 8 / 68 T-: 14 / 49 / 63 Tot: 74 / 57 / 131 Value (95% CI): Se 81.1% (72.2%–90.0%); Sp 86.0% (77.0%–95.0%); PPV 88.2% (80.6%–95.9%); NPV 77.8% (67.5%–88.0%) 3) CA-125 ≥ 35 (for all ages) Dis+ / Dis- / Tot T+: 74 / 83 / 157 T-: 16 / 171 / 187 Tot: 90 / 254 / 344 Value (95% CI): Se 82.3% (74.4%–90.2%); Sp 67.3% (61.5%–73.1%); PPV 47.1% (39.3%–54.9%); NPV 91.4% (87.4%–95.5%) 4) CA-125 ≥ 65 (for all ages)	**Comments:** --Data stratified by age/menopausal status for CA-125 using lower cutpoint not presented. --Appears that borderline tumors grouped with malignant --Unclear how patients chosen; no definition of menopause **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				Dis+ Dis- Tot T+ 68 154 222 T- 22 100 122 Tot 90 254 344 Value Lower 95% CI Upper 95% CI Se 75.6% 66.7% 84.4% Sp 39.4% 33.4% 45.4% PPV 30.6% 24.6% 36.7% NPV 82.0% 75.1% 88.8%	
Gadducci, Ferdeghini, Rispoli, et al., 1991 **#6490**	**Geographical location:** Pisa, Italy University Hospital **Dates:** NR **Size of population:** 220 women **Other** Preop patients at university hospital **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** NR but serum drawn prior to surgery	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR aside from undergoing gynecological surgery (presumably for mass) **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** None	1) CA-125 > 35 U/ml Dis+ Dis- Tot T+ 49 66 115 T- 8 97 105 Tot 57 163 220 Value Lower 95% CI Upper 95% CI Se 86.0% 77.0% 95.0% Sp 59.5% 52.0% 67.0% PPV 42.6% 33.6% 51.6% NPV 92.4% 87.3% 97.5%	**Comments:** --No description on inclusion etc. --Hospital based study --Although info in article on TATI, this was excluded from 2x2 table because it's not common test **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:+ Sample size: - not discussed Statistical tests: + Blinding: +/- not discussed but prospective? Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: CA-125 > 35 (or 65) U/ml				
Grab, Flock, Stohr, et al., 2000 **#2720**	**Geographical location:** Germany **Dates:** NR **Size of population:** 101 **Other** Case series **Reference standard:** Yes **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** NR **Statistical tests used:** None **Blinding:** Yes **Definition of positive and negative on screening test:** Sonography: Adnexal masses with suspicious patterns according to Sawai's classification, DePriest score ≥ 5 or RI < 0.45 assumed to be malignant	**Age:** Median: 45 Range: 18-82 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Asymptomatic adnexal mass Prospective consecutive patients scheduled for laproscopy offered entry **Exclusion criteria:** Pregnant women, age < 18 and patients with clinical symptoms of malignancy	**Symptomatic (n [%]):** NR – but assume 0% since excluded **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Ultrasound – combination morphology and Doppler Dis+ / Dis- / Tot T+: 11 / 36 / 47 T-: 1 / 53 / 54 Tot: 12 / 89 / 101 Value (Lower 95% CI – Upper 95% CI) Se: 92.0% (76.7% – 100.0%) Sp: 60.0% (49.8% – 70.2%) PPV: 23.4% (11.3% – 35.5%) NPV: 98.1% (94.6% – 100.0%) 2) MRI Dis+ / Dis- / Tot T+: 10 / 14 / 24 T-: 2 / 75 / 77 Tot: 12 / 89 / 101 Value (Lower 95% CI – Upper 95% CI) Se: 83.0% (61.7% – 100.0%) Sp: 84.0% (76.4% – 91.6%) PPV: 41.7% (21.9% – 61.4%) NPV: 97.4% (93.8% – 100.0%) 3) PET Dis+ / Dis- / Tot T+: 7 / 18 / 25 T-: 5 / 71 / 76 Tot: 12 / 89 / 101 Value (Lower 95% CI – Upper 95% CI)	**Comments:** --No description of who refused surgery --Unclear how patients came to have diagnosis of mass --Descriptive analysis of MRI and CT; no scoring system used --RI cut point (0.45) not described why chosen --No discussion of inter/intra observer variability --Unclear if combination morphology and Doppler used in series or parallel --One of few studies to explicitly state presence or absence of symptoms **Quality assessment:** Reference standard: Verification bias: + Test reliability/variability: + Sample size: + Statistical tests: - Blinding: + Definition of +/- on screening test: +, but how all 3 modalities used not described

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Se	**58.0%**	30.1%	85.9%
Sp	**80.0%**	71.7%	88.3%
PPV	28.0%	10.4%	45.6%
NPV	93.4%	87.8%	99.0%

MRI
Lesions considered benign if one or more of the following were met: cystic structures without any solid areas, diameter 4 cm or less wall thickness < 3 mm and presence of typical characteristics of dermoid cyst or endometrioma. If one of these not fulfilled then considered malignant.

PET
If uptake of F-FDG equaled or exceeded that of the liver and they were not localized within structures with physiologic uptake.

COMBINATION
All 3 used in conference, but criteria not described

4) All 3 (definition of positive/negative not given)

	Dis+	Dis-	Tot
T+	11	13	24
T-	1	76	77
Tot	**12**	**89**	101

	Value	Lower 95% CI	Upper 95% CI
Se	**92.0%**	76.7%	100.0%
Sp	**85.0%**	77.6%	92.4%
PPV	45.8%	25.9%	65.8%
NPV	98.7%	96.2%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Granberg, Norstrom, and Wikland, 1990 **#5320**	**Geographical location:** Sweden **Dates:** 1987-1988 **Size of population:** 180 **Other** Prospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Used the same MD for all exams **Statistical tests used:** None **Blinding:** NR **Definition of positive and negative on screening test:** "Classified as malignant the more complex it looked on ultrasound"	**Age:** Range: < 20 to > 70 **Menopausal status (n [%]):** Pre (< 45): 86 (48%) Post (> 55): 94 (52%) **Race/ethnicity (n [%]):** Swedish **Risk factors (n [%]):** NR **Inclusion criteria:** Women scheduled for elective surgery due to adnexal masses **Exclusion criteria:** NR	**Symptomatic (n [%]):** 71% had symptoms **Detected by exam (n [%]):** 100% found at a gyn exam performed 1 week to 1 month prior to surgery, but unclear whether symptoms present **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Vaginal ultrasound (data not presented by menopausal status) Dis+ / Dis- / Tot T+: 32 / 11 / 43 T-: 7 / 130 / 137 Tot: 39 / 141 / 180 Value (Lower 95% CI, Upper 95% CI) Se: 82.1% (70.0%, 94.1%) Sp: 92.2% (87.8%, 96.6%) PPV: 74.4% (61.4%, 87.5%) NPV: 94.9% (91.2%, 98.6%)	**Comments:** --No US scoring system used – descriptive only --Unclear how patients selected (if consecutive) **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: + Statistical tests: - Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Granberg, Norstrom, and Wikland, 1991 **#10920**	**Geographical location:** Sweden **Dates:** May 1988 – Dec 1988 **Size of population:** 50 **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not really **Statistical tests used:** Student's T test Linear regression Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** US – at least one of the following criteria fulfilled: 1) tumor > 10 cm in diameter (excluding simple completely unilocular cysts), 2) unilocular with echogenic areas inside the cyst, 3) multilocular with more than one thick (> 1 mm) septation and internal	**Age:** Mean: 53.8 Range: 21-92 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Surgical series **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US – all patients |	Dis+	Dis-	Tot
---	---	---	---
T+	16	6	22
T-	0	28	28
Tot	16	34	50 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	81.3%	100.0%
Sp	82.0%	69.1%	94.9%
PPV	72.7%	54.1%	91.3%
NPV	100.0%	89.3%	100.0%		**Comments:** --No scoring system used for US morphology – descriptive and not reproducible --Clinical pathway not described in patients --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	echoes, 4) multilocular-solid.				
Guerriero, Ajossa, Garau, et al., 2005 **#7470**	**Geographical location:** Cagliari, Italy **Dates:** NR **Size of population:** 424 women 453 masses **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not really **Statistical tests used:** Se, Sp Kappa statistic, **Blinding:** NR **Definition of positive and negative on screening test:** US morphology : benign was anything that resembled an endometrioma, or a cystic teratoma, or with appearance of non-malignant(not defined) Doppler – not clearly stated but appears to be	**Age:** Mean (SD): 39 (15) Range: 14-79 **Menopausal status (n [%]):** Pre (< 45): 323 (76%) Post (> 55): 101 (24%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US morphology Dis+ Dis- Tot T+ 95 64 159 T- 0 294 294 Tot 95 358 453 Value / Lower 95% CI / Upper 95% CI Se 100.0% 96.8% 100.0% Sp 82.0% 78.0% 86.0% PPV 59.7% 52.1% 67.4% NPV 100.0% 99.0% 100.0% 2) Doppler Dis+ Dis- Tot T+ 95 32 127 T- 0 326 326 Tot 95 358 453 Value / Lower 95% CI / Upper 95% CI Se 100.0% 96.8% 100.0% Sp 91.0% 88.0% 94.0% PPV 74.8% 67.3% 82.4% NPV 100.0% 99.1% 100.0%	**Comments:** --Definition of positive morphology scan or Doppler very unclear (used some subjective description) – no score or calculation used --Kappa statistic calculated --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	simple presence or absence of flow visualized in "echogenic structure"				
Guerriero, Ajossa, Risalvato, et al., 1998 **#3400**	**Geographical location:** Italy **Dates:** Jan 1996-May 1997 **Size of population:** 240 178 women with 192 masses **Other** Prospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Kappa for reliability **Blinding:** NR **Definition of positive and negative on screening test:** B-mode: Malignant when echogenic structure situated adjacent to wall of cyst is present, when a large > 3 mm irregular	**Age:** Mean (SD): 41 (15) Range: 14-77 **Menopausal status (n [%]):** Pre (< 45): 127 (71%) Post (> 55): 51 (29%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women under observation for presence of adnexal mass **Exclusion criteria:** Pregnant	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Estimates are for masses not women 1) Post menopause PI ≤ 1 Dis+ / Dis- / Tot T+: 23 / 12 / 35 T-: 3 / 13 / 16 Tot: 26 / 25 / 51 Value / Lower 95% CI / Upper 95% CI Se: 88.0% / 75.5% / 100.0% Sp: 52.0% / 32.4% / 71.6% PPV: 65.7% / 50.0% / 81.4% NPV: 81.3% / 62.1% / 100.0% 2) Total for PI ≤ 1 Dis+ / Dis- / Tot T+: 29 / 84 / 113 T-: 4 / 75 / 79 Tot: 33 / 159 / 192 Value / Lower 95% CI / Upper 95% CI Se: 88.0% / 76.9% / 99.1% Sp: 47.0% / 39.2% / 54.8% PPV: 25.7% / 17.6% / 33.7% NPV: 94.9% / 90.1% / 99.8%	**Comments:** --2x2 analysis of masses not women US – descriptive no scoring system used --Unclear why different PI cut points used --No explanation for why RI cut point chosen --Good use of kappa **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: +/- Sample size: - Statistical tests: + Blinding: +/ Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	homogeneous or heterogeneous echogenic structure present or when an irregular thickened > 3 mm wall or septum present. Color Doppler imaging RI < 0.4, PI ≤ 1 or a PI ≤ 0.8 CA-125: 35 and 65U/ml				

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Guerriero, Alcazar, Coccia, et al., 2002 **#2130**	**Geographical location:** Italy **Dates:** Apr 1997 to Jul 2000 **Size of population:** 789 women with 826 masses **Other** Case series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Z statistic K statistic for agreement **Blinding:** NR **Definition of positive and negative on screening test:** TV sonography: multiloculated, complex or solid mass in which the echo architecture was not highly indicative of a benign histologic type was categorized as malignant. Color Doppler:	**Age:** Mean (SD): 40 (14) Range: 14-81 **Menopausal status (n [%]):** Pre (< 45): 617 (78%) Post (> 55): 172 (22%) **Race/ethnicity (n [%]):** NR **Inclusion criteria:** All women scheduled for surgery in the presence of a persistent adnexal mass **Exclusion criteria:** Women with an anechoic unilocular or bilocular cystic mass with a thin regular wall without endocystic vegetation	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Transvaginal sonography: Premenopausal Dis+ / Dis- / Tot T+: 48 / 62 / 110 T-: 1 / 506 / 507 Tot: 49 / 568 / 617 Value (95% CI) Se: 98.0% (94.1% – 100.0%) Sp: 89.0% (86.4% – 91.6%) PPV: 43.6% (34.4% – 52.9%) NPV: 99.8% (99.4% – 100.0%) 2) Transvaginal sonography: Postmenopausal Dis+ / Dis- / Tot T+: 91 / 84 / 175 T-: 0 / 88 / 88 Tot: 91 / 172 / 263 Value (95% CI) Se: 100.0% (96.7% – 100.0%) Sp: 51.0% (43.5% – 58.5%) PPV: 52.0% (44.6% – 59.4%) NPV: 100.0% (96.6% – 100.0%) 3) Color Doppler – Premenopausal Dis+ / Dis- / Tot T+: 46 / 23 / 69 T-: 3 / 545 / 548 Tot: 49 / 568 / 617 Value (95% CI) Se: 94.0% (87.4% – 100.0%) Sp: 96.0% (94.4% – 97.6%) PPV: 66.7% (55.5% – 77.8%) NPV: 99.5% (98.8% – 100.0%)	**Comments:** --CA-125 used but data not presented separately --Estimates are for masses, not individuals – difficult to determine denominator being used **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: -; not detailed enough to reproduce

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	malignancy was assumed if arterial flow was visualized in an echogenic structure or in an irregular solid portion defined as malignant on B-mode imaging.			4) Color Doppler: Postmenopausal Dis+ Dis- Tot T+ 87 40 127 T- 4 132 136 Tot 91 172 263 Value Lower 95% CI Upper 95% CI Se 96.0% 92.0% 100.0% Sp 77.0% 70.7% 83.3% PPV 68.5% 60.4% 76.6% NPV 97.1% 94.2% 99.9% 5) Color Doppler: combined ages Dis+ Dis- Tot T+ 133 63 196 T- 7 677 684 Tot 140 740 880 Value Lower 95% CI Upper 95% CI Se 95.0% 91.4% 98.6% Sp 91.5% 89.5% 93.5% PPV 67.9% 61.3% 74.4% NPV 99.0% 98.2% 99.7% 6) US morphology – combined ages Dis+ Dis- Tot T+ 139 146 285 T- 1 594 595 Tot 140 740 880 Value Lower 95% CI Upper 95% CI Se 99.3% 97.9% 100.0% Sp 80.3% 77.4% 83.1% PPV 48.8% 43.0% 54.6% NPV 99.8% 99.5% 100.0%	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Hata, Hata, and Kitao, 1995 **#10960**	**Geographical location:** Japan **Dates:** NR **Size of population:** 102 **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not really **Statistical tests used:** Se, Sp, Kappa statistic Chi square **Blinding:** NR **Definition of positive and negative on screening test:** US morphology : “features that suggested the possibility of malignancy” such as dense irregular septa, multilocular cysts, papillary formation, poorly defined borders, solid focus, echogenic core R< 0.72 PSV > 16 cm/sec	**Age:** Mean: 46.1 Range: 20-78 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** Japanese **Risk factors (n [%]):** NR **Inclusion criteria:** Referred to hospital with mass who had US prior to surgical evaluation **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Peak systolic velocity > 16 cm/sec Dis+ / Dis- / Tot T+ 25 6 31 T- 5 66 71 Tot 30 72 102 Value / Lower 95% CI / Upper 95% CI Se 83.3% 70.0% 96.6% Sp 91.7% 85.3% 98.1% PPV 80.6% 66.7% 94.6% NPV 93.0% 87.0% 98.9% 2) RI <0.72 Dis+ / Dis- / Tot T+ 28 23 51 T- 2 49 51 Tot 30 72 102 Value / Lower 95% CI / Upper 95% CI Se 93.3% 84.4% 100.0% Sp 68.1% 57.3% 78.9% PPV 54.9% 41.2% 68.6% NPV 96.1% 90.8% 100.0% 3) US morphology Dis+ / Dis- / Tot T+ 26 22 48 T- 4 50 54 Tot 30 72 102 Value / Lower 95% CI / Upper 95% CI Se 86.7% 74.5% 98.9% Sp 69.4% 58.8% 80.0% PPV 54.2% 40.1% 68.3% NPV 92.6% 85.6% 99.6%	**Comments:** --US morphology descriptive – no scoring system used, --Unclear why RI of 0.72 was used or PSV of 16cm/sec --LMP tumors grouped in with malignant --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Hata, Hata, Manabe, et al., 1992 **#5010**	**Geographical location:** Japan **Dates:** NR **Size of population:** 63 **Other** Prospective series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** CA-125 – yes RI - no **Statistical tests used:** ROC curves T tests, Chi-square **Blinding:** NR **Definition of positive and negative on screening test:** CA-125 > 35 U/ml RI – calculated from own analysis of data < 0.72 US – not described MRI – malignant = size > 4 cm and (any of the following): 1) cystic, wall > 3 mm +/- nodularity 2) predom solid lesion 3) involvement of other organs or sidewalls or	**Age:** Mean: 47.4 Range: 20-78 **Menopausal status (n [%]):** Pre (< 45): 35 (56%) Post (> 55): 28 (44%) **Race/ethnicity (n [%]):** Japanese **Risk factors (n [%]):** NR **Inclusion criteria:** Suspected pelvic tumors **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI < 0.72 Dis+ / Dis- / Tot T+: 25 / 17 / 42 T-: 2 / 19 / 21 Tot: 27 / 36 / 63 Value (Lower 95% CI – Upper 95% CI) Se: 92.6% (82.7% – 100.0%) Sp: 52.8% (36.5% – 69.1%) PPV: 59.5% (44.7% – 74.4%) NPV: 90.5% (77.9% – 100.0%) 2) US Dis+ / Dis- / Tot T+: 23 / 11 / 34 T-: 4 / 25 / 29 Tot: 27 / 36 / 63 Value (Lower 95% CI – Upper 95% CI) Se: 85.2% (71.8% – 98.6%) Sp: 69.4% (54.3% – 84.5%) PPV: 67.6% (51.9% – 83.4%) NPV: 86.2% (73.7% – 98.8%) 3) MRI Dis+ / Dis- / Tot T+: 18 / 1 / 19 T-: 9 / 35 / 44 Tot: 27 / 36 / 63 Value (Lower 95% CI – Upper 95% CI) Se: 66.7% (48.9% – 84.5%) Sp: 97.1% (91.6% – 100.0%) PPV: 94.7% (84.7% – 100.0%) NPV: 79.5% (67.6% – 91.5%) 4) CA-125 > 35	**Comments:** --Unclear how patients selected --Borderline tumors grouped in with malignant --No US scoring system used (and means of diagnosis not well described) --MRI almost scoring system --RI cut point determined from analysis of this data --Unable to stratify by age **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	omental dz or ascites or adenopathy			Dis+ Dis- Tot T+ 16 3 19 T- 11 33 44 Tot 27 36 63 Value Lower 95% CI Upper 95% CI Se 59.3% 40.8% 77.8% Sp 91.7% 82.7% 100.0% PPV 84.2% 67.8% 100.0% NPV 75.0% 62.2% 87.8%	
Herrmann Jr., Locher, and Goldhirsch, 1987 **#6840**	**Geographical location:** Germany **Dates:** 1981-1985 **Size of population:** 312/404 **Screening study** Retrospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Chi-square **Blinding:** NR **Definition of positive and negative on**	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** Very young age Pregnancy Endocrinologic disorder Recurrent tumors No pathology diagnosis	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US (borderline tumors excluded) Dis+ Dis- Tot T+ 38 8 46 T- 14 177 191 Tot 52 185 237 Value Lower 95% CI Upper 95% CI Se 73.1% 61.0% 85.1% Sp 95.7% 92.7% 98.6% PPV 82.6% 71.7% 93.6% NPV 92.7% 89.0% 96.4% 2) US (borderline tumors considered benign) Dis+ Dis- Tot T+ 38 11 49 T- 14 178 192 Tot 52 189 241 Value Lower 95% CI Upper 95% CI Se 73.1% 61.0% 85.1% Sp 94.2% 90.8% 97.5% PPV 77.6% 65.9% 89.2% NPV 92.7% 89.0% 96.4%	**Comments:** --Excluded 92 patients who did not get operated on within 3 weeks of sonography – delay may be related to test result --Borderline tumors were excluded from authors calculations, but reported separately --Data reported separately for pelvic versus adnexal masses, except for benign tumors, which were reported together (Table 1). This may raise numbers in Dis- column of 2x2 table with corresponding error for PPV and NPV. --Data presented on page 779 re: prevalence of disease by age, but need additional information to fill in 2x2. --Very unclear how patients selected US scoring system from Fleischer et al (not well used criteria) – and not described in text **Quality assessment:** Reference standard: - Verification bias: - Test reliability/variability:- Sample size: + Statistical tests:+ Blinding: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	screening test: Ultrasound: Thick septae, irregular solid parts within a mass, indefinite margins, and the presence of ascites and matted bowel loops regarded as malignant.				Definition of +/- on screening test: +
Hillaby, Aslam, Salim, et al., 2004 **#1620**	**Geographical location:** London, UK **Dates:** Apr 2000 – Jun 2003 **Size of population:** 119 women **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Interobserver reliability for 2 examiners on 15 cases showed agreement for crescent sign **Statistical tests used:** Se, Sp **Blinding:** Yes (prospective study) **Definition of positive and negative on screening test:**	**Age:** Mean: 43 Range: 15-81 **Menopausal status (n [%]):** Pre: 70 (70%) Post: 30 (30%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Scheduled to undergo surgery for adnexal pathology, referred to tertiary referral gyn scanning unit **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US – ovarian crescent sign (T+ = negative) Dis+ / Dis- / Tot T+: 23 / 18 / 41 T-: 1 / 58 / 59 Tot: 24 / 76 / 100 Value (Lower 95% CI – Upper 95% CI) Se: 96.0% (88.2% – 100.0%) Sp: 76.0% (66.4% – 85.6%) PPV: 56.1% (40.9% – 71.3%) NPV: 98.3% (95.0% – 100.0%) 2) PI < 1.0 Dis+ / Dis- / Tot T+: 14 / 8 / 22 T-: 10 / 68 / 78 Tot: 24 / 76 / 100 Value (Lower 95% CI – Upper 95% CI) Se: 58.0% (38.3% – 77.7%) Sp: 89.0% (82.0% – 96.0%) PPV: 63.6% (43.5% – 83.7%) NPV: 87.2% (79.8% – 94.6%) 3) CA-125 ≥ 35 U/ml Dis+ / Dis- / Tot T+: 21 / 26 / 47 T-: 3 / 50 / 53 Tot: 24 / 76 / 100	**Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: ±, crescent sign evaluated for reliability, but only in 15 cases and 2 observers. Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Yes
Ovarian crescent sign – presence of normal ovarian tissue on TVUS: Criteria to identify normal ovarian tissues were: Hypoechogenic tissue with or without ovarian follicles located adjacent to the cyst wall, which could not be separated from the cyst by applying a moderate amount of pressure and which was enclosed within the ovarian capsule encircling the tumor.
Also
CA-125 ≥ 35 U/ml
PI < 1.0

	Value	Lower 95% CI	Upper 95% CI
Se	**88.0%**	75.0%	100.0%
Sp	**66.0%**	55.3%	76.7%
PPV	44.7%	30.5%	58.9%
NPV	94.3%	88.1%	100.0%

4) Tumor volume ≥ 180 mL

	Dis+	Dis-	Tot
T+	19	35	54
T-	5	41	46
Tot	**24**	**76**	100

	Value	Lower 95% CI	Upper 95% CI
Se	**79.0%**	62.7%	95.3%
Sp	**54.0%**	42.8%	65.2%
PPV	35.2%	22.4%	47.9%
NPV	89.1%	80.1%	98.1%

5) Papillary proliferations

	Dis+	Dis-	Tot
T+	10	10	20
T-	14	66	80
Tot	**24**	**76**	100

	Value	Lower 95% CI	Upper 95% CI
Se	**41.0%**	21.3%	60.7%
Sp	**87.0%**	79.4%	94.6%
PPV	50.0%	28.1%	71.9%
NPV	82.5%	74.2%	90.8%

6) Time-averaged maximum velocity (TAMXV ≥ 12 cm/s)

	Dis+	Dis-	Tot
T+	15	7	22
T-	9	69	78
Tot	**24**	**76**	100

		Lower	Upper

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				Value 95% CI 95% CI Se 63.0% 43.7% 82.3% Sp 91.0% 84.6% 97.4% PPV 68.2% 48.7% 87.6% NPV 88.5% 81.4% 95.6%	
Hogdall, Hogdall, Tingulstad, et al., 2000 **#2610**	**Geographical location:** Denmark **Dates:** Sep 1994 to Apr 1996 **Size of population:** 168 **Screening study** Prospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** ? Interassay coefficient? **Statistical tests used:** ROC curves Mann-Witney Spearman-Rank **Blinding:** NR **Definition of positive and negative on screening test:** Not pre-specified but present Table 3 with	**Age:** Benign Median: 48 Range: 19-86 Non-ovarian cancer Median: 69 Range:42-79 Ovarian cancer Median:61.5 Range:31-82 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** Danish? **Risk factors (n [%]):** NR **Inclusion criteria:** Presence of a pelvic mass and a decision taken to proceed with surgical exploration **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Overall sensitivity for CA-125 using a cutpoint of 35U/ml Dis+ Dis- Tot T+ 34 28 62 T- 10 96 106 Tot 44 124 168 Value Lower 95% CI Upper 95% CI Se 77.3% 64.9% 89.7% Sp 77.4% 70.0% 84.8% PPV 54.8% 42.5% 67.2% NPV 90.6% 85.0% 96.1%	**Comments:** --Data are presented in Table 4 by age 50, but total N seems to indicate that this is based on only the women with ovarian cancer; determination of specificity doesn't seem valid. **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	"generally accepted cutpoints"				
Hricak, Chen, Coakley, et al., 2000 **#2800**	**Geographical location:** San Francisco, CA University Hospital **Dates:** Apr 1993 – May 1996 **Size of population:** 128 women (187 masses) **Other** Consecutive patients referred for MRI from gynecologist who had surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes; inter- and intraobserver variability explicitly measured **Statistical tests used:** Logistic regression ROC curves Se, Sp Kappa **Blinding:** Yes **Definition of positive and negative on**	**Age:** Mean: 53 Range: 18-83 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women with diagnosis of adnexal mass referred for MR from Gynoncol clinic who subsequently underwent surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Compared gadolinium enhanced MRI versus not enhanced	1) Non-enhanced MRI Dis+ / Dis- / Tot T+: **83** / **23** / 106 T-: **13** / **68** / 81 Tot: 96 / 91 / 187 Value (Lower 95% CI – Upper 95% CI) Se: 86.5% (79.6% – 93.3%) Sp: 74.7% (65.8% – 83.7%) PPV: 78.3% (70.5% – 86.1%) NPV: 84.0% (76.0% – 91.9%) 2) Gadolinium-enhanced MRI Dis+ / Dis- / Tot T+: **91** / **19** / 110 T-: **5** / **72** / 77 Tot: 96 / 91 / 187 Value (Lower 95% CI – Upper 95% CI) Se: 94.8% (90.3% – 99.2%) Sp: 79.1% (70.8% – 87.5%) PPV: 82.7% (75.7% – 89.8%) NPV: 93.5% (88.0% – 99.0%)	**Comments:** --LMP tumors grouped into malignant --Referral population from gynecological clinic - (probability of malignancy before imaging 51%); sicker population, not representative --Data collected and analyzed per mass, not per patient **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - not discussed Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	screening test: [From ref 19] Malignant = at least one of the following primary criteria present: > 4 cm, bilateral, predominantly solid, cystic with wall or septum > 3 mm or papillary projections. OR at least 2 of the following secondary criteria present: ascites, peritoneal metastasis, adenopathy.				
Huber, Medl, Baumann, et al., 2002 **#5700**	**Geographical location:** Austria **Dates:** May 1995 – Jan 2001 **Size of population:** 93 **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Chi square Fisher exact test **Blinding:** Yes **Definition of positive and negative on**	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients suspected of having ovarian cancer in time frame referred for surgery and had imaging done **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US morphology Dis+ / Dis- / Tot T+: 54 / 8 / 62 T-: 9 / 22 / 31 Tot: 63 / 30 / 93 Value (95% CI) Se: 85.0% (76.2% – 93.8%) Sp: 73.0% (57.1% – 88.9%) PPV: 87.1% (78.8% – 95.4%) NPV: 71.0% (55.0% – 86.9%) 2) MRI Dis+ / Dis- / Tot T+: 56 / 4 / 60 T-: 7 / 26 / 33 Tot: 63 / 30 / 93 Value (95% CI) Se: 89.0% (81.3% – 96.7%) Sp: 86.0% (73.6% – 98.4%) PPV: 93.3% (87.0% – 99.6%) NPV: 78.8% (64.8% – 92.7%)	**Comments:** --Patients all referred with suspicion of ovarian cancer (hence high incidence of cancer in this group) --Unclear and not reproducible criteria for + or - US and MRI – no scoring system used --Combination TVUS and abdominal US (unable to stratify, no N stated for each) **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	screening test: US – "detection of complex (noncystic) and/or solid mass, which was ≥ 5 cm in premenopausal woman or any size in postmenopausal woman. MRI – descriptive				
Hurteau, Woolas, Jacobs, et al., 1995 **#4060**	**Geographical location:** Patients from London, UK **Dates:** NR **Size of population:** Unclear – article mentions 100 patients preop evaluation as well as 88 "healthy subjects", but analysis done on 92 **Other** Series in single center **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Sen, Sp Student's t test **Blinding:** NR – prospective sampling	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Unclear – preop with diagnosis of adnexal mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 35 Dis+ / Dis- / Tot T+: 32 / 31 / 63 T-: 7 / 30 / 37 Tot: 39 / 61 / 100 Value / Lower 95% CI / Upper 95% CI Se: 82.1% / 70.0% / 94.1% Sp: 49.2% / 36.6% / 61.7% PPV: 50.8% / 38.4% / 63.1% NPV: 81.1% / 68.5% / 93.7%	**Comments:** --Very unclear patient selection, inclusion and exclusion criteria (numbers don't match up – no explanation of how went from 100 to 92) --Data on IL 2 alpha not included in 2x2 table as this is not a common test --Inclusion of healthy subjects not necessarily appropriate for diagnostic (as opposed to screening) test **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + - not discussed in article but well established test Sample size: - Statistical tests: +/- Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: CA-125 > 35 U/mL				
Inoue, Fujita, Nakazawa, et al., 1992 **#5120**	**Geographical location:** Osaka, Japan University Hospital **Dates:** Sep 1989 – May 1991 **Size of population:** 382 women **Other** Patents who underwent surgery for adnexal mass **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** ROC curves Se, Sp **Blinding:** No **Definition of positive and negative on screening test:** CA-125 > 65 U/mL CEA > 2.4 ng/mL	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women undergoing surgery for (presumed) adnexal mass at one of the University hospitals in time frame **Exclusion criteria:** LMP tumors other than those of surface epithelial-stromal type and non-gynecological tumors.	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** None	1) CA-125 (> 65 U/mL) Dis+ / Dis- / Tot T+: 25 / 44 / 69 T-: 40 / 273 / 313 Tot: 65 / 317 / 382 Value (95% CI) Se: 38.0% (26.2% – 49.8%) Sp: 86.0% (82.2% – 89.8%) PPV: 36.2% (24.9% – 47.6%) NPV: 87.2% (83.5% – 90.9%) 2) CEA Dis+ / Dis- / Tot T+: 14 / 10 / 24 T-: 51 / 307 / 358 Tot: 65 / 317 / 382 Value (95% CI) Se: 22.0% (11.9% – 32.1%) Sp: 97.0% (95.1% – 98.9%) PPV: 58.3% (38.6% – 78.1%) NPV: 85.8% (82.1% – 89.4%) Additional data reported for other markers, Sialyl-Tn (STN), sialys-Lewis Xi (SLX), CA 19-9, Tissue polypeptide antigen	**Comments:** --CA-125 limit 65U/mL --No description of patient population at all --5 surface epithelial tumors of LMP were grouped into malignant category **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + (COV discussed in serum samples between labs) Sample size: - (not discussed) Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Itakura, Kikkawa, Kajiyama, et al., 2003 **#1690**	**Geographical location:** Japan University Hospital **Dates:** Jun 1998 – Jul 2000 **Size of population:** 84 women (95 tumors) **Other** Hospital referral for surgery secondary to mass **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** CA-125 – yes DePriest – yes PI - yes **Statistical tests used:** Se, Sp **Blinding:** Prospective study – blinding not discussed **Definition of positive and negative on screening test:** CA-125 > 65 U/mL Morphological index of DePriest score > 7 PI < 1.0	**Age:** Mean (SD): 49.1 **Menopausal status (n [%]):** Pre: 41 (48.8%) Post : 43 (51.2%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients who underwent surgery at university hospital for mass (not clearly described) **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Morphological index of DePriest (> 7) |	Dis+	Dis-	Tot
---	---	---	---
T+	24	15	39
T-	3	42	45
Tot	27	57	84 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	90.3%	79.1%	100.0%
Sp	73.4%	61.9%	84.9%
PPV	61.5%	46.3%	76.8%
NPV	93.3%	86.0%	100.0% | 2) PI (min < 1.0) |	Dis+	Dis-	Tot
---	---	---	---
T+	22	8	30
T-	5	49	54
Tot	27	57	84 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	80.6%	65.7%	95.5%
Sp	85.9%	76.9%	94.9%
PPV	73.3%	57.5%	89.2%
NPV	90.7%	83.0%	98.5% | 3) CA-125 (> 65 U/ml) |	Dis+	Dis-	Tot
---	---	---	---
T+	19	7	26
T-	8	50	58
Tot	27	57	84 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	70.4%	53.2%	87.6%
Sp	87.7%	79.2%	96.2%
PPV	73.1%	56.0%	90.1%
NPV	86.2%	77.3%	95.1%		**Comments:** --Se and Sp reported unclear if for patient or for tumor (most likely for tumor) --CA-125 cutoff 65 U/mL --Borderline tumors lumped into malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - not discussed Statistical tests: +/- Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Jacobs, Oram, Fairbanks, et al., 1990 **#6820**	**Geographical location:** London **Dates:** NR **Size of population:** 143 **Other** Consecutive series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** CA-125 - yes **Statistical tests used:** Stepwise logistic regression Se, Sp RMI **Blinding:** NR – but prospective **Definition of positive and negative on screening test:** CA-125 > 30U/mL US: 1 point assigned for each of the following: --multilocular cyst --solid areas --metastases --ascites --bilateral lesion	**Age:** Mean (SD): Benign 48.8 (14.3) Malignant 59.0 (11.8) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Consecutive admissions for elective surgical investigation of pelvic mass in hospital in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 30 |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 34 | 25 | 59 | | T- | 8 | 76 | 84 | | Tot | 42 | 101 | 143 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 81.0% | 69.1% | 92.9% | | Sp | 75.0% | 66.6% | 83.4% | | PPV | 57.6% | 45.0% | 70.2% | | NPV | 90.5% | 84.2% | 96.8% | 2) US score ≥ 1 |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 41 | 52 | 93 | | T- | 0 | 46 | 46 | | Tot | 41 | 98 | 139 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 100.0% | 92.7% | 100.0% | | Sp | 46.9% | 37.0% | 56.8% | | PPV | 44.1% | 34.0% | 54.2% | | NPV | 100.0% | 93.5% | 100.0% | 3) US score ≥ 2 |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 29 | 17 | 46 | | T- | 12 | 81 | 93 | | Tot | 41 | 98 | 139 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 70.7% | 56.8% | 84.6% | | Sp | 82.7% | 75.2% | 90.2% | | PPV | 63.0% | 49.1% | 77.0% | | NPV | 87.1% | 80.3% | 93.9% |	**Comments:** --CA-125 cutoff of 30U/mL used **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: + (but didn't explain why 30 U/mL used as cutoff for CA-125)

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Jacobs, Stabile, Bridges, et al., 1988 **#6830**	**Geographical location:** London, UK **Dates:** NR **Size of population:** 1010 screened with BME and CA-125 58 US done secondary to abnormal BME or CA-125 9 went to surgery **Screening study** **Reference standard:** Surgery for 9 repeat US and BME at 3 month intervals for one year for initial abnormals (by CA-125 or BME) **Reference standard applied to all test negatives?:** No **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** BME – "palpable mass of any size . . . separate from uterus or GI tract" CA-125 ≥ 30 U/ml US ovarian volume > 8.8ml	**Age:** Mean: 54 Range: 45-83 **Menopausal status (n [%]):** Post (> 55): 1010 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 30 U/ml (assuming all test negatives truly negative) |	Dis+	Dis-	Tot
---	---	---	---
T+	1	30	31
T-	0	979	979
Tot	1	1009	1010 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	-	100.0%
Sp	97.0%	96.0%	98.1%
PPV	3.2%	0.0%	9.4%
NPV	100.0%	99.7%	100.0% | 2) CA-125 > 30U/ml not assuming test negatives true negative (including only those which had US) |	Dis+	Dis-	Tot
---	---	---	---
T+	1	30	31
T-	0	27	27
Tot	1	57	58 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	-	100.0%
Sp	47.4%	34.4%	60.3%
PPV	3.2%	0.0%	9.4%
NPV	100.0%	88.9%	100.0% | 3) BME (assuming all negatives true negative) |	Dis+	Dis-	Tot
---	---	---	---
T+	1	27	28
T-	0	982	982
Tot	1	1009	1010 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
[illegible]	[illegible]	-	100.0%		**Comments:** --Unclear where definition of ovarian volume as abnormal (> 8.8ml) came from --Vague criteria for BME --Low incidence of cancer in this screening study --Abdominal US only **Quality assessment:** Reference standard: - Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

			200.0%	
Sp	97.3%	96.3%		98.3%
PPV	3.6%	0.0%		10.4%
NPV	100.0%	99.7%		100.0%

4) BME not assuming test negatives true negative

	Dis+	Dis-	Tot
T+	1	27	28
T-	0	30	30
Tot	1	57	58

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	- 200.0%	100.0%
Sp	52.6%	39.7%	65.6%
PPV	3.6%	0.0%	10.4%
NPV	100.0%	90.0%	100.0%

4) US alone (ovarian volume > 8.8ml)

	Dis+	Dis-	Tot
T+	1	12	13
T-	0	45	45
Tot	1	57	58

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	- 200.0%	100.0%
Sp	78.9%	68.4%	89.5%
PPV	7.7%	0.0%	22.2%
NPV	100.0%	93.3%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Jain, 1994 **#4620**	**Geographical location:** Stanford, CA **Dates:** NR **Size of population:** 42 women (50 masses) **Other** Prospective series of surgical cases with US **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** RI – yes US – not really (references his own article the next in this batch #4950) **Statistical tests used:** NR **Blinding:** NR – but prospective **Definition of positive and negative on screening test:** RI < 0.4 Grey scale US – Presence of any of the following: Irregular solid portion, irregular wall, thick irregular septa, mural notdule; Doppler done,	**Age:** Mean: 43 Range: 33-55 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** 42 women with clinically suspected adnexal masses undergoing surgery – US performed 1-5 days prior **Exclusion criteria:** Obstetrical cases	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI < 0.4 of MASSES Dis+ / Dis- / Tot T+: 7 / 7 / 14 T-: 2 / 33 / 35 Tot: 9 / 40 / 49 Value (Lower 95% CI – Upper 95% CI) Se: 77.8% (50.6% – 100.0%) Sp: 82.5% (70.7% – 94.3%) PPV: 50.0% (23.8% – 76.2%) NPV: 94.3% (86.6% – 100.0%) 2) US (MASSES) Dis+ / Dis- / Tot T+: 9 / 2 / 11 T-: 0 / 38 / 38 Tot: 9 / 40 / 49 Value (Lower 95% CI – Upper 95% CI) Se: 100.0% (66.7% – 100.0%) Sp: 95.0% (88.2% – 100.0%) PPV: 81.8% (59.0% – 100.0%) NPV: 100.0% (92.1% – 100.0%)	**Comments:** --Data presented such that LMP could be easily excluded from analysis --Unclear if anyone other than author did US examinations --No discussion of inter/intra operator variability --Data presented with N = masses not patients **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: - Blinding: +/- Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	but neovascularization not required for diagnosis of malignancy				
Jain, Friedman, Pettinger, et al., 1993 **#4950**	**Geographical location:** Davis and Palo Alto, CA **Dates:** NR **Size of population:** 32 **Other** Prospective series **Reference standard:** Histopathology or FNA at time of laparoscopy **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** MRI - **Statistical tests used:** Kappa, Se, Sp **Blinding:** NR – but prospective **Definition of positive and negative on screening test:** Descriptive diagnostic criteria for both MRI and US – "simple cyst, hem cyst, endometrioma, dermoid, pedunc fibroid, ovarian carcinoma, ovarian torsion"	**Age:** Mean: 41.5 Range: 29-54 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Prospective women with suspected masses at hospital scheduled for surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US: Cancer vs. benign Dis+ / Dis- / Tot T+: 3 / 14 / 17 T-: 0 / 20 / 20 Tot: 3 / 34 / 37 Value (Lower 95% CI – Upper 95% CI) Se: 100.0% (0.0% – 100.0%) Sp: 60.0% (43.5% – 76.5%) PPV: 17.6% (0.0% – 35.8%) NPV: 100.0% (85.0% – 100.0%) 2) MRI: Cancer vs. benign Dis+ / Dis- / Tot T+: 2 / 0 / 2 T-: 1 / 34 / 35 Tot: 3 / 34 / 37 Value (Lower 95% CI – Upper 95% CI) Se: 67.0% (13.8% – 100.0%) Sp: 100.0% (91.2% – 100.0%) PPV: 100.0% (0% – 100.0%) NPV: 97.1% (91.6% – 100.0%)	**Comments:** --Kappa calculated --Unclear how individuals chosen for study (if not consecutive) --US score consisted of diagnosis as did MRI – ?reproducible --Pathology not available for all , laproscopy patients had FNA with examination of ovaries --Se/Sp based on masses, not patients **Quality assessment:** Reference standard: + Verification bias: +/- Test reliability/variability: + Sample size: - Statistical tests: +/- Blinding: + Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Juhasz, Kurjak, Lampe, et al., 1990 **#10860**	**Geographical location:** Yugoslavia and Hungary **Dates:** NR **Size of population:** 147 **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** Presence or absence of color flow within adnexal mass on Doppler	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Color flow present or absent |	Dis+	Dis-	Tot
---	---	---	---
T+	21	17	38
T-	1	108	109
Tot	22	125	147 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	95.5%	86.8%	100.0%
Sp	86.4%	80.4%	92.4%
PPV	55.3%	39.5%	71.1%
NPV	99.1%	97.3%	100.0%		**Comments:** --Unclear exactly what was meant by present flow on Doppler --No description of patient characteristics --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Kawahara, Yoshida, Kurokawa, et al., 2004 **#10**	**Geographical location:** Fukui, Japan **Dates:** Sept 2001 – Aug 2003 **Size of population:** 38 **Other** Series of suspected ovarian cancer cases who went to surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** MRI – yes PET scan - ? **Statistical tests used:** Se, Sp **Blinding:** NR – prospective study **Definition of positive and negative on screening test:** MRI – if any of these features was met, mass considered suspicious for malignancy: "cystic without solid areas, diameter of 4cm or less, wall thickness < 0.3 cm, the presence of typical characteristics of	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients who had been screened in gynecological oncology clinic with BME and US and considered to have masses suspicious for malignancy **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) MRI Dis+ Dis- Tot T+ 21 2 23 T- 2 13 15 Tot 23 15 38 Value / Lower 95% CI / Upper 95% CI Se 91.3% 79.8% 100.0% Sp 86.7% 69.5% 100.0% PPV 91.3% 79.8% 100.0% NPV 86.7% 69.5% 100.0% 2) FDG-PET Dis+ Dis- Tot T+ 18 0 18 T- 5 15 20 Tot 23 15 38 Value / Lower 95% CI / Upper 95% CI Se 78.3% 61.4% 95.1% Sp 100.0% 80.0% 100.0% PPV 100.0% 83.3% 100.0% NPV 75.0% 56.0% 94.0%	**Comments:** --LMP tumors grouped in with malignant --Patient referral from onc clinic, not representative, sicker --MRI scoring system vague and not reproducible --No discussion of inter/intra observer variability --In reporting PET results authors state "the benign tumors were correctly identified as negative for malignancy in all 13 patients with benign lesion" however there were 15 patients with benign lesions **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: - Blinding: +/- Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	dermoid cyst or endometrioma" For PET scan with 18-Flourodeoxyglucose – "hypermetabolic lesions that were more intense than the liver and not attributable to bladder etc. were considered positive for malignancy"				
Kawai, Kikkawa, Ishikawa, et al., 1994 **#10940**	**Geographical location:** Japan **Dates:** Apr 1990 – Aug 1993 **Size of population:** 109 **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Student's t test Chi square analysis **Blinding:** NR **Definition of positive and negative on screening test:**	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Doppler Dis+ / Dis- / Tot T+: 30 / 14 / 44 T-: 10 / 55 / 65 Tot: 40 / 69 / 109 Value (Lower 95% CI – Upper 95% CI) Se: 75.0% (61.6% – 88.4%) Sp: 79.2% (69.6% – 88.8%) PPV: 68.2% (54.4% – 81.9%) NPV: 84.6% (75.8% – 93.4%) 2) CA-125 > 35U/ml Dis+ / Dis- / Tot T+: 29 / 37 / 66 T-: 11 / 32 / 43 Tot: 40 / 69 / 109 Value (Lower 95% CI – Upper 95% CI) Se: 72.2% (58.3% – 86.1%) Sp: 45.8% (34.0% – 57.6%) PPV: 43.9% (32.0% – 55.9%) NPV: 74.4% (61.4% – 87.5%) 3) CA-72-4 Dis+ / Dis- / Tot T+: 17 / 12 / 29 T-: 23 / 57 / 80	**Comments:** --Means of evaluating Doppler (1/PI) is unusual and not justified in text --Cut point for PI of 1.25 is also unusual --LMP tumors grouped in with malignant --Calculated PPV and NPV differ slightly from text --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Doppler: 1/PI with cutoff 0.8 (which equals a cutoff for PI of 1.25)

CA-125 > 35 U/ml
CA-72-4 > 4U/ml
CA-19-9 > 37U/ml

Tot	40	69	109

	Value	Lower 95% CI	Upper 95% CI
Se	41.7%	26.4%	57.0%
Sp	83.3%	74.5%	92.1%
PPV	58.6%	40.7%	76.5%
NPV	71.3%	61.3%	81.2%

4) CA-19-9

	Dis+	Dis-	Tot
T+	14	12	26
T-	26	57	83
Tot	40	69	109

	Value	Lower 95% CI	Upper 95% CI
Se	36.1%	21.2%	51.0%
Sp	83.3%	74.5%	92.1%
PPV	53.8%	34.7%	73.0%
NPV	68.7%	58.7%	78.7%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Komatsu, Konishi, Mandai, et al., 1996 **#4050**	**Geographical location:** Kyoto, Japan **Dates:** May 1989 – May 1993 **Size of population:** 82 **Other** Retrospective case series comparing US and MR **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** **Statistical tests used:** **Blinding:** Yes **Definition of positive and negative on screening test:** US classification: 0 = cyst with well-defined, thin wall 1a = septation 1b = solid tissue 2a = complex mass with internal structure with diffuse low-level echoes, no distinct findings of cyst 2b = complex mass with internal structure with	**Age:** Mean: 45.9 Range: 17-89 **Menopausal status (n [%]):** Pre (< 45): 54 (65.9%) Post (> 55): 28 (34.1%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US: benign vs. malignant (US class 0 or 1a benign, all other malignant); borderline counted as malignancy Dis+ / Dis- / Tot T+: 34 / 26 / 60 T-: 0 / 22 / 22 Tot: 34 / 48 / 82 Value / Lower 95% CI / Upper 95% CI Se: 100.0% / 91.2% / 100.0% Sp: 45.8% / 31.7% / 59.9% PPV: 56.7% / 44.1% / 69.2% NPV: 100.0% / 86.4% / 100.0% 2) MRI: benign vs. malignant (MRI class 1b malignant, all others benign); borderline counted as malignancy Dis+ / Dis- / Tot T+: 31 / 3 / 34 T-: 3 / 22 / 25 Tot: 34 / 25 / 59 Value / Lower 95% CI / Upper 95% CI Se: 91.2% / 81.6% / 100.0% Sp: 88.0% / 75.3% / 100.0% PPV: 91.2% / 81.6% / 100.0% NPV: 88.0% / 75.3% / 100.0%	**Comments:** --Unclear how patients chosen for study (consecutive? . . .) --Clinical presentation not described --Outcome of 73 patients who did not undergo surgery not described --Over half of masses were malignant --Results not stratified by age/menopausal status **Quality assessment:** Reference standard: +; pathology Verification bias: -; large portion did not undergo surgery Test reliability/variability:-; not described Sample size:-; wide CIs Statistical tests: -; 2x2 tables not presented Blinding: + Definition of +/- on screening test:

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	excrescences with low-level echoes and blurred margin 1b, 2a, 2b considered possibly malignant MRI: 0 = cyst with well-defined wall, no internal enhancement, homogenous low intensity on T-1 weighted images, high intensity on T2 1a = neoplasm with septation 1b = solid 2a = cystic mass with complex fluid such as blood, proteinaceous fluid, or fat 2b = cystic mass with unenhanced excrescences 1b considered possibly malignant				

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Kurjak and Kupesic, 1999 **#2920**	**Geographical location:** Zagreb, Croatia University Hospital **Dates:** Jan 1997 – Jun 1998 **Size of population:** 120 women **Other** Patients scheduled for surgery at university hospital **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Se, Sp **Blinding:** No – but prospective **Definition of positive and negative on screening test:** TVUS score ≥ 5 (where +2 for papillarities > 3 mm, +1 for shadowing present, +1 for septa > 3 mm, +2 for solid parts present, +2 for mixed or high echogenicity, +1 for peritoneal fluid, +2 for RI ≤ 0.42	**Age:** Mean (SD): Pre 34 Peri 49 Post 61 Range: 18-77 **Menopausal status (n [%]):** Pre: 76 (63.3%) Peri: 7 (5.8%) Post: 37 (30.8%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Not clearly stated – women with masses to undergo surgery in hospital Premenopausal women had US during early proliferative phase only **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) 2D US combined |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 10 | 3 | 13 | | T- | 1 | 106 | 107 | | Tot | 11 | 109 | 120 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 90.9% | 73.9% | 100.0% | | Sp | 97.3% | 94.2% | 100.0% | | PPV | 76.9% | 54.0% | 99.8% | | NPV | 99.1% | 97.2% | 100.0% | 2) 3D US combined |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 11 | 1 | 12 | | T- | 0 | 108 | 108 | | Tot | 11 | 109 | 120 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 100.0% | 72.7% | 100.0% | | Sp | 99.1% | 97.3% | 100.0% | | PPV | 91.7% | 76.0% | 100.0% | | NPV | 100.0% | 97.2% | 100.0% |	**Comments:** --TVUS and 3D/Doppler scoring systems invented by authors (?) and no reference of use given (no reliability calculation etc.) – reproducibility? **--Same group of patients as #2820** **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:- Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	3D US score > 7 (where +2 for papilarities, +1 for present shadowing, +1 for septa >3 mm, +2 for solid parts present, +2 for mixed or high echogenicity, +1 for peritoneal fluid present, +2 for irregular surface, +2 for relation with surrounding structures disturbed, +2 for chaotic vessel arrangement, +2 for complex branching pattern)				
Kurjak, Kupesic, Sparac, et al., 2000 **#2560**	**Geographical location:** Zagreb, Croatia University Hospital **Dates:** Jan 1998 – Jun 1999 **Size of population:** 90 women **Other** Prospective patients with masses for surgery in university hospital **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No (although same scale used in articles #2820 or #2920)	**Age:** Mean (SD): Pre – 34 Peri – 49 Post – NR Median: Range: 18-77 **Menopausal status (n [%]):** Pre: 58 (64.4%) Peri: 4 (4.4%) Post: 28 (31.1%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Prospective patients with ovarian mass scheduled to Have surgery in hospital **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) 2D alone (B-mode) |	Dis+	Dis-	Tot
---	---	---	---
T+	6	5	11
T-	3	76	79
Tot	9	81	90 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	66.7%	35.9%	97.5%
Sp	93.8%	88.6%	99.1%
PPV	54.5%	25.1%	84.0%
NPV	96.2%	92.0%	100.0% | 2) 2D Doppler alone RI ≤ 0.42 |	Dis+	Dis-	Tot
---	---	---	---
T+	8	4	12
T-	1	77	78
Tot	9	81	90 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	88.9%	68.4%	100.0%
Sp	95.1%	90.3%	99.8%
PPV	66.7%	40.0%	93.3%
NPV	98.7%	96.2%	100.0%		**Comments:** --The scale used for scoring invented by authors – not independently verified, not part of the literature (used also in #2920) --Differentiation of US from Doppler assessment in terms of scale not clear **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: +/- Blinding: + Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Statistical tests used:
Se, Sp

Blinding:
Yes

Definition of positive and negative on screening test:
2D US
for 2D alone score ≥ 3 is positive
for Doppler alone score ≥ 2 is positive
for combined 2D score ≥ 5 score is positive where +2 for papillarities, +1 for shadowing present, +1 for septa > 3 mm thick, +2 for solid parts present, +2 for mixed of high level echogenicity, +1 for peritoneal fluid present, +2 for RI ≤ 0.42

3D US score ≥ 5 is+
for Doppler alone score ≥ 2+ for combined score ≥ 7 where +2 for papillarities, +1 for shadowing present, +1 for septa > 3 mm thick, +2 for solid parts present, +2 for mixed of high level echogenicity, +1 for peritoneal fluid present, +2 for irregular surface, +2 for disturbed relation with surrounding structures, +2 for chaotic vessel arrangement, +2 for complex branching pattern

3) Combined 2D US and Doppler

	Dis+	Dis-	Tot
T+	8	2	10
T-	1	79	80
Tot	9	81	90

	Value	Lower 95% CI	Upper 95% CI
Se	88.9%	68.4%	100.0%
Sp	97.5%	94.1%	100.0%
PPV	80.0%	55.2%	100.0%
NPV	98.8%	96.3%	100.0%

4) 3D TVUS

	Dis+	Dis-	Tot
T+	7	2	9
T-	2	79	81
Tot	9	81	90

	Value	Lower 95% CI	Upper 95% CI
Se	77.8%	50.6%	100.0%
Sp	97.5%	94.1%	100.0%
PPV	77.8%	50.6%	100.0%
NPV	97.5%	94.2%	100.0%

5) 3D Doppler

	Dis+	Dis-	Tot
T+	8	2	10
T-	1	79	80
Tot	9	81	90

	Value	Lower 95% CI	Upper 95% CI
Se	88.9%	68.4%	100.0%
Sp	97.5%	94.1%	100.0%
PPV	80.0%	55.2%	100.0%
NPV	98.8%	96.3%	100.0%

6) Combined 3D

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	9	1	10
T-	0	80	80
Tot	9	81	90

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	66.7%	100.0%
Sp	98.8%	96.4%	100.0%
PPV	90.0%	71.4%	100.0%
NPV	100.0%	96.3%	100.0%

Kurjak and Predanic, 1992

#4990

Geographical location:
Zagreb Croatia
University Hospital

Dates: Sep 1990 – Sep 1991

Size of population:
812 women screened with US in whom 174 masses detected in 154 women

Other
Combination – initially screening of women with "clinical suspicion of mass" then analysis of subset who went to surgery (n = 154)

Reference standard:
For subset - Histopathology

Reference standard applied to all test negatives?:
Yes for subset

Test reliability

Age:
Mean (SD): 48
Range: 19-76

Menopausal status (n [%]):
Pre: 111 (72%)
Post: 43 (28%)

Race/ethnicity (n [%]):
NR

Risk factors (n [%]):
NR

Inclusion criteria:
Initially, all women referred to hospital in time frame with clinical suspicion of adnexal mass. Then those who had mass on US and went to surgery

Exclusion criteria:
NR

Symptomatic (n [%]):
NR

Detected by exam (n [%]):
Presumably 100%

Detected by imaging (n [%]):
Not applicable

Combination (n [%]):
NR

Additional data used for diagnosis:
NR

1) Morphologic scoring system

	Dis+	Dis-	Tot
T+	35	7	42
T-	3	129	132
Tot	38	136	174

	Value	Lower 95% CI	Upper 95% CI
Se	92.1%	83.5%	100.0%
Sp	94.8%	91.1%	98.5%
PPV	83.3%	72.1%	94.6%
NPV	97.7%	95.2%	100.0%

2) Color Doppler scoring system

	Dis+	Dis-	Tot
T+	37	0	37
T-	1	136	137
Tot	38	136	174

	Value	Lower 95% CI	Upper 95% CI
Se	97.3%	92.1%	100.0%
Sp	100.0%	97.8%	100.0%
PPV	100.0%	91.9%	100.0%
NPV	99.3%	97.8%	100.0%

3) Combined Doppler and morphology

Comments:
--Article attempts to verify scoring system these authors developed and used previously (in #2820 and #2560) (modification of Sassone criteria)
--Data analyzed in terms of masses not individuals

Quality assessment:
Reference standard: +
Verification bias: +
Test reliability/variability: -
Sample size: -
Statistical tests: - (no discussion in any of their papers of kappa etc.)
Blinding: +/-
Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	established?: No **Statistical tests used:** Se, Sp **Blinding:** NR but prospective **Definition of positive and negative on screening test:** TVUS for 2D alone score ≥ 3 is positive for Doppler alone score ≥ 2 is positive for combined 2D score ≥ 5 score is positive where +2 for papillarities, +1 for shadowing present, +1 for septa > 3 mm thick, +2 for solid parts present, +2 for mixed of high level echogenicity, +1 for peritoneal fluid present, +2 for RI ≤ 0.42			Dis+ / Dis- / Tot T+: 37 / 0 / 37 T-: 1 / 136 / 137 Tot: 38 / 136 / 174 Value (Lower 95% CI – Upper 95% CI) Se: 97.3% (92.1% – 100.0%) Sp: 100.0% (97.8% – 100.0%) PPV: 100.0% (91.9% – 100.0%) NPV: 99.3% (97.8% – 100.0%)	
Kurjak, Schulman, Sosic, et al., 1992 **#5020**	**Geographical location:** Zagreb, Croatia **Dates:** 1989-1990 **Size of population:** 1000 screened 83 operated on **Screening study** For N = 83, operative series **Reference standard:** Histopathology (for N = 83)	**Age:** NR **Menopausal status (n [%]):** Post (> 55): 83 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age > 40 At least 12 months since	**Symptomatic (n [%]):** For N = 1000, 257 (25.7%) were symptomatic For N = 83 of the 29 with malignant tumors, 25 were symptomatics **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** For N = 1000, 83 had US that lead to surgery (8.3%) **Combination (n [%]):**	1) For RI < 0.41 Dis+ / Dis- / Tot T+: 28 / 3 / 31 T-: 1 / 51 / 52 Tot: 29 / 54 / 83 Value (Lower 95% CI – Upper 95% CI) Se: 96.0% (88.9% – 100.0%) Sp: 95.0% (89.2% – 100.0%) PPV: 90.3% (79.9% – 100.0%) NPV: 98.1% (94.3% – 101.0%) 2) US morphology	**Comments:** --US scoring system although illustrated in figure 1, was not described in text – unclear where and how scoring system derived, if tested/verified . . . --RI cutoff of < 0.41 used based upon distribution of RI data points in from this study itself – also similar to prior retrospective cutpoint by study authors --Doppler Index – although descriptive in nature, appears in data to be identical to the RI results, with the exception that the RI "unable to measure" are included in results as 0

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Reference standard applied to all test negatives?: Yes for N = 83 **Test reliability established?:** Yes **Statistical tests used:** Chi-square Fisher exact Student t test Se, Sp **Blinding:** NR – but prospective **Definition of positive and negative on screening test:** RI < 0.41 US morphology score > 4 (figure 1): Color Doppler score – 0 for vessels not seen, 1 for regular separate vessels, 2 for randomly dispersed vessels	LMP Criteria for surgery included: Mass 5 cm or greater on at least 2 exams Cyst associated with persistent or acute pain Enlarged ovary or cyst with RI "near" 0.4 on 2 separate **Exclusion criteria:** NR	NR **Additional data used for diagnosis:** NR	Dis+ Dis- Tot T+ 14 1 15 T- 15 53 68 Tot 29 54 83 Value Lower 95% CI Upper 95% CI Se 48.0% 29.8% 66.2% Sp 98.0% 94.3% 100.0% PPV 93.3% 80.7% 100.0% NPV 77.9% 68.1% 87.8% 3) For Doppler score (> 2) Dis+ Dis- Tot T+ 26 1 27 T- 3 53 56 Tot 29 54 83 Value Lower 95% CI Upper 95% CI Se 89.7% 78.6% 100.0% Sp 98.1% 94.6% 100.0% PPV 96.3% 89.2% 100.0% NPV 94.6% 88.7% 100.0% 4) For combined US and Doppler (< 6) Dis+ Dis- Tot T+ 26 3 29 T- 3 51 54 Tot 29 54 83 Value Lower 95% CI Upper 95% CI Se 89.7% 78.6% 100.0% Sp 94.4% 88.3% 100.0% PPV 89.7% 78.6% 100.0% NPV 94.4% 88.3% 100.0%	leading to a drop in sensitivity of results --No discussion of inter/intra observer variability --Numbers in Table 4 inconsistent --Followup of initial test negatives not reported **Quality assessment:** Reference standard: + Verification bias: + for surgical series, - for screening series Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - - but prospective Definition of +/- on screening test: + for RI, +/- for US

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Kurjak, Shalan, Kupesic, et al., 1994 **#4470**	**Geographical location:** Zagreb, Croatia **Dates:** Jan 1988 to Dec 1992 **Size of population:** 5013 screened 38 operated on **Screening study** **Reference standard:** Histopathology for few who went to surgery Otherwise repeat US **Reference standard applied to all test negatives?:** No But US repeated in initial abnormals (cycstic structures less than 5cm) followed up after 6 months **Test reliability established?:** Yes **Statistical tests used:** NR - proportions **Blinding:** NR - prospective **Definition of positive and negative on screening test:** "a persistently enlarged ovary" – 2.5 cm or greater in two separate scans RI < 0.41	**Age:** Mean (SD): Pre 45.1 (3.5) Post 56.2 (5.2) Range: 4-71 **Menopausal status (n [%]):** Pre (< 45): 2214 (44%) Post (> 55): 2799 (56%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age ≥ 40 No "pelvic symptoms" **Exclusion criteria:** Women on hormonal therapy	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** 0 (0%) **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Premenopausal women scanned during day 3-8 of cycle **STUDY FLOW –** 5013 screened 424 abnormal ovaries, of whom 20 went to surgery, leaving 404 repeat US (316 still pending), of whom 70 resolved spontaneously 18 persistent went to surgery	1) US – persistent mass; only those with surgical confirmation |	Dis+	Dis-	Tot
---	---	---	---
T+	4	1	5
T-	0	27	27
Tot	4	28	32 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	25.0%	100.0%
Sp	96.4%	89.6%	100.0%
PPV	80.0%	44.9%	100.0%
NPV	100.0%	88.9%	100.0% | 2) US, assuming all test negative true negatives, excluding 316 with results not reported |	Dis+	Dis-	Tot
---	---	---	---
T+	4	1	5
T-	0	97	97
Tot	4	98	102 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	25.0%	100.0%
Sp	99.0%	97.0%	100.0%
PPV	80.0%	44.9%	100.0%
NPV	100.0%	96.9%	100.0%		**Comments:** --Decision to operate not described --Screening series not complete – 316 women undergoing followup US still (from total of 404 needing it!) – this is confusing --Unclear what was used in US diagnosis (assume from title combination of doppler and US morphology, but nothing in article) --US followup after 6 months, but span of time not mentioned --Assume 100% followup? – no discussion of drop out etc. --No discussion of inter/intra observer variability --Results not stratified by age/menopausal status **Quality assessment:** Reference standard: +/- Verification bias: -- Test reliability/variability: - Sample size: - Statistical tests: - (no significance testing done) Blinding: + Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	US morphology score > 4				
Kurjak, Zalud, and Alfirevic, 1991 **#5190**	**Geographical location:** Zagreb, Croatia **Dates:** NR **Size of population:** 14317 total 8,620 asymptomatic for screening 5697 with "suspected adnexal mass" 680 operated on **Screening study** **Reference standard:** Histopathology in 680 operated on **Reference standard applied to all test negatives?:** Only to surgical cases **Test reliability established?:** Yes **Statistical tests used:** Not described - proportions **Blinding:** Not mentioned (prospective) **Definition of positive and negative on screening test:** RI < 0.4	**Age:** Mean: Pre 42 Post 56 Range: 18-72 **Menopausal status (n [%]):** Pre (< 45): 7495 Post (> 55): 1125 This is of the 8620 women referred from clinic **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** Unclear **Detected by exam (n [%]):** ?5697/14,317: **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI < 0.4 Dis+ Dis- Tot T+ 54 1 55 T- 2 623 625 Tot 56 624 680 Value / Lower 95% CI / Upper 95% CI Se 96.4% 91.6% 100.0% Sp 99.8% 99.5% 100.0% PPV 98.2% 94.7% 100.0% NPV 99.7% 99.2% 100.0%	**Comments:** --No mention of inclusion criteria for this study --8620 were asymptomatic screening, 5697 were referral from gyn clinic for "suspicion" of mass (not stated how this diagnosis made), however, both grouped into analysis together --Proportion cited pre and post menopausal don't add up to total N, rather to "clinic referral" group --No mention of how many RI they were unable to measure --No mention of followup --Very problematic study with large numbers . . . --RI mean calculated from 5 separate readings (not the lowest measured as in other studies) --Inter observer variability WAS discussed --RI < 0.4 used here, same authors use 0.41 elsewhere --Unclear how much overlap in study sample between the authors different papers there is --16/55 malignancies Stage I **Quality assessment:** Reference standard: +/- Verification bias: - Test reliability/variability: + Sample size: - Statistical tests:- Blinding: +/- Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Kurtz, Tsimikas, Tempany, et al., 1999 **#2940**	**Geographical location:** Ann Arbor, MI Baltimore, MD Boston, MA Philadelphia, PA University Hospitals **Dates:** May 1993 – Apr 1996 **Size of population:** 280 women **Other** All eligible patients referred by Gyn Onc to University hospital for surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Se, Sp, ROC **Blinding:** Yes to outcome (prospective) radiologists not blinded to diagnosis **Definition of positive and negative on screening test:**	**Age:** Mean: 52 Median: 51.5 Range 19-82 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Over 18, suspected of having ovarian cancer based on physical exam or pelvic US **Exclusion criteria:** Unable to provide consent, not a surgical candidate, pregnancy, prior surgery within 6 months of entry into study	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Doppler and Conventional US Dis+ / Dis- / Tot T+ 82 / 6 / 88 T- 27 / 149 / 176 Tot 109 / 155 / 264 Value (Lower 95% CI – Upper 95% CI) Se 75.2% (67.1% – 83.3%) Sp 96.1% (93.1% – 99.2%) PPV 93.2% (87.9% – 98.4%) NPV 84.7% (79.3% – 90.0%) 2) CT Dis+ / Dis- / Tot T+ 58 / 16 / 74 T- 5 / 134 / 139 Tot 63 / 150 / 213 Value (Lower 95% CI – Upper 95% CI) Se 92.1% (85.4% – 98.7%) Sp 89.3% (84.4% – 94.3%) PPV 78.4% (69.0% – 87.8%) NPV 96.4% (93.3% – 99.5%) 3) MRI Dis+ / Dis- / Tot T+ 47 / 16 / 63 T- 1 / 115 / 116 Tot 48 / 131 / 179 Value (Lower 95% CI – Upper 95% CI) Se 97.9% (93.9% – 100.0%) Sp 87.8% (82.2% – 93.4%) PPV 74.6% (63.9% – 85.4%) NPV 99.1% (97.5% – 100.0%)	**Comments:** --Data in 2x2 tables derived from ROC curves, estimated based on total N's for each test; numbers agree with Table 6 in manuscript. --Referral base for study from oncology clinic – sicker pop **Quality assessment:** Reference standard: + Verification bias: +, few women were excluded Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Leeners, Schild, Funk, et al., 1996 **#3940**	**Geographical location:** Aachen, Germany Academic **Dates:** Jan 1993 – Sep 1994 **Size of population:** 101 patients (109 tumors) 95 women got surgery **Other** Consecutive series in single center **Reference standard:** Histopathology for 95 Unclear what for the other 6 **Reference standard applied to all test negatives?:** No – see above **Test reliability established?:** Yes **Statistical tests used:** Fisher exact Wilcoxon 2 sample test **Blinding:** NR but prospective **Definition of positive and negative on screening test:** Sassone's score > 9 PI < 0.65 RI < 0.45	**Age:** Mean: 48.4 Range: 16-84 **Menopausal status (n [%]):** Pre (< 45): 67 (66.3%) Post (> 55): 34(33.7%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Family history: 16 (15.8%) **Inclusion criteria:** Consecutive patients referred to US for "clinical suspicion of an adnexal mass" **Exclusion criteria:** NR	**Symptomatic (n [%]):** 52 (51.5%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Sassone score Dis+ / Dis- / Tot T+: 16 / 13 / 29 T-: 7 / 73 / 80 Tot: 23 / 86 / 109 Value (Lower 95% CI – Upper 95% CI) Se: 69.8% (51.0% – 88.5%) Sp: 85.0% (77.5% – 92.5%) PPV: 55.2% (37.1% – 73.3%) NPV: 91.3% (85.1% – 97.4%) 2) Doppler (RI – lowest from a series of measurements) Dis+ / Dis- / Tot T+: 16 / 34 / 50 T-: 7 / 52 / 59 Tot: 23 / 86 / 109 Value (Lower 95% CI – Upper 95% CI) Se: 70.9% (52.3% – 89.5%) Sp: 60.9% (50.6% – 71.2%) PPV: 32.0% (19.1% – 44.9%) NPV: 88.1% (79.9% – 96.4%) 3) Combined Doppler and Sassone score Dis+ / Dis- / Tot T+: 17 / 23 / 40 T-: 6 / 63 / 69 Tot: 23 / 86 / 109 Value (Lower 95% CI – Upper 95% CI) Se: 74.0% (56.1% – 91.9%) Sp: 73.7% (64.4% – 83.0%) PPV: 42.5% (27.2% – 57.8%) NPV: 91.3% (84.7% – 98.0%)	**Comments:** --Borderline tumors grouped in with malignant --2x2 tables calculated in terms of masses not patients --Patients had been referred from gynecologic clinic where had often times already had US --No discussion of followup of 6 initial who didn't get surgery --RI and PI cutoff calculated from data itself – not using prior cutoffs --Clinical presentation not described **Quality assessment:** Reference standard: +/- Verification bias: +/- Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Lerner, Timor-Tritsch, Federman, et al., 1994 **#6360**	**Geographical location:** New York, NY University Hospital **Dates:** May 1990 – Mar 1993 **Size of population:** 312 patients with 350 ovarian masses **Other** Retrospective analysis of US of women who had gone to surgery for adnexal mass **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes for Sasonne, no for modified criteria used here **Statistical tests used:** Se, Sp **Blinding:** Yes **Definition of positive and negative on screening test:** Modified Sasonne where + was ≥ 3 and +1 for shadowing present, +1 for ≥ 3 mm walls, +2 for solid wall structure, +3	**Age:** Mean: 44.5 Range: 12-85 **Menopausal status (n [%]):** Pre (< 45): 228 (73%) Post (> 55): 84 (27%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women who had surgery in time frame for whom images were available **Exclusion criteria:** Attempts made to not perform US in luteal phase	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US scoring system of Sassone (cutoff ≥ 3 for T+) |	Dis+	Dis-	Tot
---	---	---	---
T+	30	72	102
T-	1	247	248
Tot	31	319	350 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	96.8%	90.6%	100.0%
Sp	77.4%	72.8%	82.0%
PPV	29.4%	20.6%	38.3%
NPV	99.6%	98.8%	100.0%		**Comments:** --The article uses data from its institution to fit a linear model from which they modify Sassonne's criteria -- their criteria actually performs worse that the original from Sassone (Se 100, Sp 83) which isn't discussed fully --Denominator in 2x2 tables is masses not individuals --LMP tumors included as benign **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: -, cutoff may have been selected *a posteriori* Sample size: - Statistical tests: -, analysis based on tumors not patients. Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	for papillarities ≥ 3 mm, +3 for mixed or high echogenicity				
Lin, Angel, DuBeshter, et al., 1993 **#4890**	**Geographical location:** Rochester, NY, USA **Dates:** Jul 1989 – Jun 1990 **Size of population:** 80 women **Other** Case series Retrospective **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No for US and CT (no scoring system used) **Statistical tests used:** P value calculated by NR **Blinding:** No; retrospective **Definition of positive and negative on screening test:** US and CT – "presence of a complex or heterogenous mass, ascites, omental tumor or other evidence of metastatic tumor in the	**Age:** Median: 56 Range: 19-88 **Menopausal status (n [%]):** Pre (< 45): 18 (22.5%) Post (> 55): 62 (77.5) **Race/ethnicity (n [%]):** Caucasian 72 (90%) Black 8 (10%) **Risk factors (n [%]):** Family history: 11 (13.8%) **Inclusion criteria:** "Mass in the pelvic area" who underwent surgery in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** Pain - 37(46.3%) Asymptomatic – 70(87.5%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US |	Dis+	Dis-	Tot
---	---	---	---
T+	15	11	26
T-	3	11	14
Tot	18	22	40 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	83.3%	66.1%	100.0%
Sp	50.0%	29.1%	70.9%
PPV	57.7%	38.7%	76.7%
NPV	78.6%	57.1%	100.0% | 2) CT |	Dis+	Dis-	Tot
---	---	---	---
T+	25	9	34
T-	4	5	9
Tot	29	14	43 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	86.2%	73.7%	98.8%
Sp	35.7%	10.6%	60.8%
PPV	73.5%	58.7%	88.4%
NPV	55.6%	23.1%	88.0%		**Comments:** --Patients referred from gynecological onconlogy clinic --No scoring system for US or CT used --Retrospective with subjective means of judging "suspicious for malignancy" and no mention of blinding or how this assessment of prior radiology studies was made --Borderline tumors grouped in with malignant --Not all tests available for all patients – hence difference in N for each 2x2 table --The PPV and NPV of CT 2x2 table differ significantly in my calculation than that reported in the article (PPV-75%, NPV-71%) **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	pelvic area or abdomen"				
Lin, Wu, Lee, et al., 1993 **#6990**	**Geographical location:** Taiwan University Hospital **Dates:** Jul 1990 – Oct 1993 **Size of population:** 370 women **Other** Hospital referrals for adnexal masses **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** RI – yes (usually with 0.4 cutoff) **Statistical tests used:** Se, Sp **Blinding:** NR but study prospective **Definition of positive and negative on screening test:** RI ≤ 0.4 or 0.5 or 0.6	**Age:** Mean: 40.5 Range: 11-81 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Suspected ovarian mass on US referred to hospital for surgery **Exclusion criteria:** Study US in luteal phase	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 370 (100%) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI ≤ 0.4 Dis+ / Dis- / Tot T+: 62 / 8 / 70 T-: 28 / 272 / 300 Tot: 90 / 280 / 370 Value (95% CI) Se: 68.9% (59.3% – 78.5%) Sp: 97.1% (95.1% – 99.1%) PPV: 88.6% (81.1% – 96.0%) NPV: 90.7% (87.4% – 94.0%) 2) RI ≤ 0.5 Dis+ / Dis- / Tot T+: 71 / 22 / 93 T-: 19 / 258 / 277 Tot: 90 / 280 / 370 Value (95% CI) Se: 78.9% (70.5% – 87.3%) Sp: 92.1% (88.9% – 95.3%) PPV: 76.3% (67.7% – 85.0%) NPV: 93.1% (90.2% – 96.1%) 3) RI ≤ 0.6 Dis+ / Dis- / Tot T+: 82 / 38 / 120 T-: 8 / 242 / 250 Tot: 90 / 280 / 370 Value (95% CI) Se: 91.1% (85.2% – 97.0%) Sp: 86.4% (82.4% – 90.4%) PPV: 68.3% (60.0% – 76.7%) NPV: 96.8% (94.6% – 99.0%)	**Comments:** --Borderline tumors grouped in malignant category --"Satisfactory arterial waveforms" only in 111(40.7%) of benign masses (and in 87(96.7%) of malignant) – however, all were included in the 2x2 table (assume 60% non-satisfactory wave forms used in benign lesions?) --Report treats no satisfactory wave form as "test negative." **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - (not discussed or calculated for operators in study) Sample size: - Statistical tests: + Blinding: +/- Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Luxman, Bergman, Sagi, et al., 1991 **#6530**	**Geographical location:** Tel Aviv, Israel **Dates:** NR **Size of population:** 102 **Other** Case series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Se, Sp **Blinding:** No **Definition of positive and negative on screening test:** US – unclear "simple" if lesion unilocular and lacking septa "complex" if solid area, papillae, septa, enhanced echogenicity	**Age:** Mean: 62 Range: 42-90 **Menopausal status (n [%]):** Post (> 55): 102 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR – presumable, presence of mass scheduled for surgery during time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US—Size > 5 cm and/or complex/solid = malignant Dis+ / Dis- / Tot T+: 27 / 42 / 69 T-: 2 / 31 / 33 Tot: 29 / 73 / 102 Value / Lower 95% CI / Upper 95% CI Se: 93.0% / 83.7% / 100.0% Sp: 42.0% / 30.7% / 53.3% PPV: 39.1% / 27.6% / 50.6% NPV: 93.9% / 85.8% / 100.0% 2) US—size > 5 cm alone = malignant Dis+ / Dis- / Tot T+: 24 / 12 / 36 T-: 5 / 61 / 66 Tot: 29 / 73 / 102 Value / Lower 95% CI / Upper 95% CI Se: 82.8% / 69.0% / 96.5% Sp: 83.6% / 75.1% / 92.1% PPV: 66.7% / 51.3% / 82.1% NPV: 92.4% / 86.0% / 98.8% 3) US—complex or solid = malignant Dis+ / Dis- / Tot T+: 27 / 33 / 60 T-: 2 / 40 / 42 Tot: 29 / 73 / 102 Value / Lower 95% CI / Upper 95% CI Se: 93.1% / 83.9% / 100.0% Sp: 54.8% / 43.4% / 66.2% PPV: 45.0% / 32.4% / 57.6% NPV: 95.2% / 88.8% / 100.0%	**Comments:** --All postmenopausal but age range 42-90 --Very unclear how tests were graded + or - (more than "simple" vs. "complex"?) **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:- Sample size: - Statistical tests: +/- Blinding: + Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Ma, Shen, and Lang, 2003 **#1900**	**Geographical location:** Peking, China University **Dates:** Jan 1998 – June 1999 **Size of population:** 140 women **Other** Retrospective analysis of surgical patients with CA-125 and US in single academic center **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** CA-125 – yes US – unclear what scoring system used – reliability probably not established **Statistical tests used:** Se, Sp, PPV **Blinding:** NR (retrospective analysis of already collected data) **Definition of positive and negative on screening test:** CA-125 > (various cutoffs analyzed 30, 50, 100)	**Age:** Range: 30 - NR **Menopausal status (n [%]):** Pre: 89 (64%) Post: 51 (36%) > 1 year of amenorrhea or if s/p hysterectomy, age > 50 years **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** "Ovarian neoplasm" patients over 30 years admitted to a single institution **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 ≥ 30 U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	56	27	83
T-	7	50	57
Tot	63	77	140 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	88.9%	81.1%	96.7%
Sp	64.9%	54.2%	75.6%
PPV	67.5%	57.4%	77.5%
NPV	87.7%	79.2%	96.2% | 2) CA-125 ≥ 50 |	Dis+	Dis-	Tot
---	---	---	---
T+	53	21	74
T-	10	56	66
Tot	63	77	140 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	84.1%	75.1%	93.1%
Sp	72.7%	62.7%	82.7%
PPV	71.6%	61.3%	81.9%
NPV	84.8%	76.2%	93.5% | 3) US score ≥ 2 |	Dis+	Dis-	Tot
---	---	---	---
T+	59	13	72
T-	4	64	68
Tot	63	77	140 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	93.7%	87.7%	99.7%
Sp	83.1%	74.7%	91.5%
PPV	81.9%	73.1%	90.8%
NPV	94.1%	88.5%	99.7%		**Comments:** --Menopause defined --Unclear where US scoring system comes from and how calculated. --Unclear if US score done at time of imaging study or when looking back. --CA-125 cutoff (30, 50) not what used in States now **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + for CA-125 - for US Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	US score 0 or 1 point given for: multi-focal lesion, nature of focal lesion, unilateral or bilateral lesion, ascites, metastasis Max = 5				
Maggino, Gadducci, D'Addario, et al., 1994 **#4500**	**Geographical location:** Padua, Pisa, Bari, Brescia, and Milan, Italy **Dates:** Mar 1991-Mar 1992 **Size of population:** 383; 48 excluded based on criteria, 45 not reported because ultrasound and CA-125 did not lead to surgery **Other** Multicenter series **Reference standard:** Surgery **Reference standard applied to all test negatives?:** No **Test reliability established?:** Not referenced or discussed (ultrasound) **Statistical tests used:** Se, Sp **Blinding:** No **Definition of positive and negative on**	**Age:** Range: 40-91 Overall mean not reported **Menopausal status (n [%]):** 100% post menopausal **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Clinical diagnosis of pelvic mass Postmenopausal at least 1 year **Exclusion criteria:** Premenopausal, Previous malignancy, except breast ca Previous bilateral adnexectomy Previous hysterectomy if < 55 years	**Symptomatic (n [%]):** 209 (72.1%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125, threshold >35, EXCLUDING 45 patients not operated on because of US and CA-125 Dis+ / Dis- / Tot T+: 83 / 24 / 107 T-: 23 / 110 / 133 Tot: 106 / 134 / 240 Value (Lower 95% CI – Upper 95% CI) Se: 78.3% (70.5% – 86.1%) Sp: 82.1% (75.6% – 88.6%) PPV: 77.6% (69.7% – 85.5%) NPV: 82.7% (76.3% – 89.1%) 2) CA-125, threshold > 65, EXCLUDING 45 patients not operated on because of US and CA-125 Dis+ / Dis- / Tot T+: 76 / 10 / 86 T-: 30 / 124 / 154 Tot: 106 / 134 / 240 Value (Lower 95% CI – Upper 95% CI) Se: 71.7% (63.1% – 80.3%) Sp: 92.5% (88.1% – 97.0%) PPV: 88.4% (81.6% – 95.1%) NPV: 80.5% (74.3% – 86.8%) 3) Ultrasound, equivocal or higher as positive, EXCLUDING patients not operated on because of findings	**Comments:** --Reference standard not applied to all test negatives -2x2 tables are limited to patients with adnexal masses (excluding other non-ovarian pelvic masses) **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

screening test:
Probably benign:

- < 5 cm
- thin, clear wall
- hypoechogenic
- no septae, or no more than 3 thin septae
- No fluid in cul de sac

Equivocal:

- Between 5 and 10 cm
- Thick, clear, smooth wall
- Hypoechogenic liquid or solid homogeneous content
- > 3 thin septae
- Thick, regular septae
- No vegetations
- No free peritoneal fluid

Malignant: none of the above

	Dis+	Dis-	Tot
T+	105	72	177
T-	1	62	63
Tot	106	134	240

	Value	Lower 95% CI	Upper 95% CI
Se	99.1%	97.2%	100.0%
Sp	46.3%	37.8%	54.7%
PPV	59.3%	52.1%	66.6%
NPV	98.4%	95.3%	100.0%

4) Ultrasound, malignant as positive, EXCLUDING patients not operated on because of findings

	Dis+	Dis-	Tot
T+	90	30	120
T-	16	104	120
Tot	106	134	240

	Value	Lower 95% CI	Upper 95% CI
Se	84.9%	78.1%	91.7%
Sp	77.6%	70.6%	84.7%
PPV	75.0%	67.3%	82.7%
NPV	86.7%	80.6%	92.7%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Malkasian Jr., Knapp, Lavin, et al., 1988 **#6810**	**Geographical location:** Rochester, MN; Boston, MA; Hershey, PA; Los Angeles, CA **Dates:** NR **Size of population:** 172; 14 excluded for total of 158 **Other** Multicenter, case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se/Sp **Blinding:** Yes **Definition of positive and negative on screening test:** CA-15 at various thresholds	**Age:** Benign Mean: 43.5 Range: 15-88 Malignant Mean: 63.5 Range: 16-96 **Menopausal status (n [%]):** Benign Pre (< 45): 56 (62.2%) Post (> 55): 34 (37.8%) Malignant: Pre 10 (14.7%) Post: 58 (85.3%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Palpable mass Scheduled for surgery Blood drawn within 1 week of surgery **Exclusion criteria:** Preop definitive diagnosis of ovarian cancer (n = 11) Blood > 1 week (n = 3)	**Symptomatic (n [%]):** Clinical presentation not described **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Stage I 23.3% II 10.0% III 61.5% IV 5.0%	1) All patients, CA-125 > 35 as positive Dis+ / Dis- / Tot T+: 53 / 18 / 71 T-: 15 / 72 / 87 Tot: 68 / 90 / 158 Value (Lower 95% CI – Upper 95% CI) Se: 77.9% (68.1% – 87.8%) Sp: 80.0% (71.7% – 88.3%) PPV: 74.6% (64.5% – 84.8%) NPV: 82.8% (74.8% – 90.7%) 2) All patients, CA-125 > 100 as threshold Dis+ / Dis- / Tot T+: 50 / 4 / 54 T-: 18 / 86 / 104 Tot: 68 / 90 / 158 Value (Lower 95% CI – Upper 95% CI) Se: 73.5% (63.0% – 84.0%) Sp: 95.6% (91.3% – 99.8%) PPV: 92.6% (85.6% – 99.6%) NPV: 82.7% (75.4% – 90.0%) 3) Premenopausal patients, CA-125 > 35 as threshold Dis+ / Dis- / Tot T+: 6 / 15 / 21 T-: 4 / 41 / 45 Tot: 10 / 56 / 66 Value (Lower 95% CI – Upper 95% CI) Se: 60.0% (29.6% – 90.4%) Sp: 73.2% (61.6% – 84.8%) PPV: 28.6% (9.2% – 47.9%) NPV: 91.1% (82.8% – 99.4%) 4) Premenopausal patients, CA-125 > 100	**Comments:** --Unclear how subjects selected --Spectrum of disease described --Tests all drawn within 1 week of surgery; unclear if results would have been different if drawn prior to decision for surgery -- Borderline tumors included with malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:+ Sample size: + (confidence intervals given) Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

as threshold

	Dis+	Dis-	Tot
T+	6	3	9
T-	4	53	57
Tot	10	56	66

	Value	Lower 95% CI	Upper 95% CI
Se	60.0%	29.6%	90.4%
Sp	94.6%	88.7%	100.0%
PPV	66.7%	35.9%	97.5%
NPV	93.0%	86.4%	99.6%

5) Postmenopausal patients, CA-125 > 35 as threshold

	Dis+	Dis-	Tot
T+	47	3	50
T-	11	31	42
Tot	58	34	92

	Value	Lower 95% CI	Upper 95% CI
Se	81.0%	70.9%	91.1%
Sp	91.2%	81.6%	100.0%
PPV	94.0%	87.4%	100.0%
NPV	73.8%	60.5%	87.1%

6) Postmenopausal patients, CA-125 > 100 as threshold

	Dis+	Dis-	Tot
T+	44	1	45
T-	13	33	46
Tot	57	34	91

	Value	Lower 95% CI	Upper 95% CI
Se	77.2%	66.3%	88.1%
Sp	97.1%	91.4%	100.0%
PPV	97.8%	93.5%	100.0%
NPV	71.7%	58.7%	84.8%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Maly, Riss, and Deutinger, 1995 **#6800**	**Geographical location:** Vienna, Austria **Dates:** NR **Size of population:** 102 women **Other** Case series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Chi-square, t-test **Blinding:** NR **Definition of positive and negative on screening test:** Diastolic notch: "short drop of flow curve at beginning of diastole"	**Age:** Range: 28-75 **Menopausal status (n [%]):** Pre (< 45): 55 (53.9%) Post (> 55): 47 (46.1%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** Clinical presentation not described **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Premenopausal US in secretory phase	1) Diastolic notch absent |	Dis+	Dis-	Tot
---	---	---	---
T+	37	5	42
T-	0	39	39
Tot	37	44	81 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	91.9%	100.0%
Sp	88.6%	79.3%	98.0%
PPV	88.1%	78.3%	97.9%
NPV	100.0%	92.3%	100.0% | 2) Demonstrable blood vessels |	Dis+	Dis-	Tot
---	---	---	---
T+	37	44	81
T-	2	19	21
Tot	39	63	102 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	94.9%	87.9%	100.0%
Sp	30.2%	18.8%	41.5%
PPV	45.7%	34.8%	56.5%
NPV	90.5%	77.9%	100.0%		**Comments:** --Unclear how patient pop chosen, if consecutive, if any excluded **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: - (not clearly defined)

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Mancuso, De Vivo, Triolo, et al., 2004 **#1610**	**Geographical location:** Italy University Hospital **Dates:** NR **Size of population:** 125 women **Other** Patients referred to hospital with mass who had surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** CA-125 - yes US – unclear what was used **Statistical tests used:** Se, Sp, LR **Blinding:** NR (US and serum prior to surgery) **Definition of positive and negative on screening test:** CA-125 ≥ 35 U/ml US – NR what was + or negative or what scoring system used	**Age:** Mean (SD): 42.2 Range: 18 - 82 **Menopausal status (n [%]):** Pre 76 (61%) Post 49 (39%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients referred to hospital with mass who had surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** 68 (54.4%) symptomatic 5 (4%) had urinary or intestinal symptoms only 30 (24%) described as asymptomatic **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** 22 (17.6%) reported a menstrual disorder as main symptom	1) US Dis+ / Dis- / Tot T+: 14 / 24 / 38 T-: 0 / 87 / 87 Tot: 14 / 111 / 125 Value (95% CI) Se: 100.0% (78.6% – 100.0%) Sp: 78.6% (71.0% – 86.2%) PPV: 36.8% (21.5% – 52.2%) NPV: 100.0% (96.6% – 100.0%) 2) CA-125 ≥ 35 U/ml Dis+ / Dis- / Tot T+: 14 / 24 / 38 T-: 0 / 87 / 87 Tot: 14 / 111 / 125 Value (95% CI) Se: 100.0% (78.6% – 100.0%) Sp: 78.6% (71.0% – 86.2%) PPV: 36.8% (21.5% – 52.2%) NPV: 100.0% (96.6% – 100.0%)	**Comments:** --Menopause versus fertile not defined --Even though data on menopausal status collected, analysis used age > or < 50 as --US scoring system not described (?Sasonne or modified) – positive or negative US not defined **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + for CA-125 - for US

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Manjunath, Pratap-kumar, Sujatha, et al., 2001 **#2510**	**Geographical location:** Manipal, India **Dates:** Jan 1997 – Aug 1999 **Size of population:** 152 women **Other** Retrospective analysis of women admitted to academic hospital with pelvic mass who had surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** CA-125 – yes US - ? **Statistical tests used:** Se, Sp, ROC curves **Blinding:** NR **Definition of positive and negative on screening test:** CA-125 (multiple cutoffs) US score 1 point given for presence of multi-locular systic lesion, solid area, bilateral, ascites, intraabdominal mets.	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 84(55.2%) Post (> 55): 64(42.1%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients who had surgery for pelvic masses **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 ≥ 35 U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	77	10	87
T-	16	45	61
Tot	93	55	148 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	83.0%	75.4%	90.6%
Sp	82.0%	71.8%	92.2%
PPV	88.5%	81.8%	95.2%
NPV	73.8%	62.7%	84.8% | 2) US score ≥ 2 |	Dis+	Dis-	Tot
---	---	---	---
T+	40	6	46
T-	53	49	102
Tot	93	55	148 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	43.0%	32.9%	53.1%
Sp	89.0%	80.7%	97.3%
PPV	87.0%	77.2%	96.7%
NPV	48.0%	38.3%	57.7%		**Comments:** --LMP tumors grouped into malignant --Although menopausal status was reported, results were not stratified by menopausal status (or age) --Unclear US scoring system **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + for CA-125 ?/- for US Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Marchetti, Zambon, Lamaina, et al., 2002 **#2230**	**Geographical location:** Padua, Italy **Dates:** Sep 1996-Oct 2001 **Size of population:** 176 positives/4,350 exams **Screening study** **Reference standard:** Surgery, followup **Reference standard applied to all test negatives?:** No **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Chi-square, Se/Sp **Blinding:** Yes **Definition of positive and negative on screening test:** Criteria for referral or positive test not described	**Age:** Mean: 49 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Borderline classified as malignant **Exclusion criteria:** NR	**Symptomatic (n [%]):** 518 (11.9%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Screening frequency not described	1) All patients with positive ultrasound, borderline classified as malignant, assuming negative ultrasound truly negative Dis+ / Dis- / Tot T+: 2 / 174 / 176 T-: 0 / 4174 / 4174 Tot: 2 / 4348 / 4350 Value (Lower 95% CI, Upper 95% CI) Se: 100.0% (0.0%, 100.0%) Sp: 96.0% (95.4%, 96.6%) PPV: 1.1% (0.0%, 2.7%) NPV: 100.0% (99.9%, 100.0%) 2) Patients undergoing surgery only (not clear from paper, but 29 had “ultrasound findings indicative of malignant lesions”, and 45 total underwent surgery Dis+ / Dis- / Tot T+: 2 / 27 / 29 T-: 0 / 16 / 16 Tot: 2 / 43 / 45 Value (Lower 95% CI, Upper 95% CI) Se: 100.0% (0.0%, 100.0%) Sp: 37.2% (22.8%, 51.7%) PPV: 6.9% (0.0%, 16.1%) NPV: 100.0% (81.3%, 100.0%)	**Comments:** --Variable reference standard --Length of followup, loss to followup not clearly described --Results not stratified by age, menopausal status **Quality assessment:** Reference standard:- Verification bias: - Test reliability/variability: - Sample size: + Statistical tests:- Blinding: + Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Marret, Sauget, Giraudeau, et al., 2004 **#7680**	**Geographical location:** France **Dates:** Feb 2002 – Mar 2003 **Size of population:** 99 women 101 masses **Prospective series ("pilot" per authors)** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Kappa statistics AUC Se, Sp Chi square – Fisher exact **Blinding:** NR **Definition of positive and negative on screening test:** RI < 0.53 CA-125 > 25 U/ml	**Age:** Mean: 46.2 Range: 19-72 **Menopausal status (n [%]):** Pre (< 45): 58 (58.6%) Post (> 55): 41 (41.4%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Woman with diagnosis of adnexal mass admitted to hospital in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI < 0.53 |	Dis+	Dis-	Tot
---	---	---	---
T+	18	13	31
T-	5	65	70
Tot	23	78	101 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	78.0%	61.1%	94.9%
Sp	83.0%	74.7%	91.3%
PPV	58.1%	40.7%	75.4%
NPV	92.9%	86.8%	98.9% | 2) CA-125 ≥ 25 |	Dis+	Dis-	Tot
---	---	---	---
T+	19	12	31
T-	4	66	70
Tot	23	78	101 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	83.0%	67.6%	98.4%
Sp	85.0%	77.1%	92.9%
PPV	61.3%	44.1%	78.4%
NPV	94.3%	88.8%	99.7%		**Comments:** --Analysis done for masses not women --Study looks specifically at the use of IV contrast at the time of US (as this is a novel method, data from that outcome is not included in this evidence table) --Unable to stratify by menopausal status --Interobserver correlation coefficient 0.92 --US modality (TVUS vs. abdominal) not specified – assume TV? **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:+ Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Matthes, Moreira de Andrade, and Bighetti, 1996 **#11020**	**Geographical location:** Brazil **Dates:** Feb 1992 – Feb 1994 **Size of population:** 51, however results only available for 43 **Reference standard:** **Reference standard applied to all test negatives?:** **Test reliability established?:** **Statistical tests used:** **Blinding:** **Definition of positive and negative on screening test:**	**Age:** Mean: 45.75 Median: 46 Range: 15-83 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Unclear **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US (Kurjak morphology) Dis+ / Dis- / Tot T+: 8 / 7 / 15 T-: 2 / 26 / 28 Tot: 10 / 33 / 43 Value / Lower 95% CI / Upper 95% CI Se: 80.0% / 55.2% / 100.0% Sp: 78.8% / 64.8% / 92.7% PPV: 53.3% / 28.1% / 78.6% NPV: 92.9% / 83.3% / 100.0%	**Comments:** --8 patients dropped from analysis – not mentioned why --Kurjak criteria for morphologic classification (without Doppler), and cutpoint not described for this study --LMP included with malignant --Unclear if TVUS or abdominal US or combination used **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
McIntosh, Drescher, Karlan, et al., 2004 **#6700**	**Geographical location:** Seattle, WA **Dates:** NR **Size of population:** 95/315 total including healthy controls (no mass) 52 ovarian cancer 43 benign ovarian tumors = 95 used for calculations here **Other** Case-control **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** ROC, LR models **Blinding:** NR **Definition of positive and negative on screening test:** SA 125 – cutoffs not selected a priori	**Age:** NR **Menopausal status (n [%]):** Pre: 90 (29%) Post: 225 (71%) **Race/ethnicity (n [%]):** White: 268 (85%) Hispanic: 4 (1%%) Black: 1 (< 1%) Asian: 8 (3%) Native American: 3 (< 1%) **Risk factors (n [%]):** NR **Inclusion criteria:** Cases and controls – randomly selected from repository **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 (cutoff not specified; sensitivity calculated corresponding to 98% specificity on ROC curve based on LR model) Dis+ / Dis- / Tot T+: 19 / 1 / 20 T-: 33 / 42 / 75 Tot: 52 / 43 / 95 Value / Lower 95% CI / Upper 95% CI Se: 37.2% / 24.1% / 50.3% Sp: 98.0% / 93.8% / 100.0% PPV: 95.0% / 85.4% / 100.0% NPV: 56.0% / 44.8% / 67.2% Also reported for soluble mesothelin related (SMR) marker; not reported here.	**Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Medl, Kulenkampff, Stiskal, et al., 1995 **#6300**	**Geographical location:** Vienna, Austria **Dates:** NR **Size of population:** 73 women **Other** Case series **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not described or discussed **Statistical tests used:** Se/Sp **Blinding:** Not described **Definition of positive and negative on screening test:** Ultrasound: • Cyst with irregular wall • Wall thickness > 3 cm • Papillary wall structures • Solid components • Presence of ascites MRI: • > 4cm	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Stage I: n=11 (29.7%) II: n=1 (2.7%) III: n=24 (64.9%) IV: n=1 (2.7%)	1) Ultrasound |	Dis+	Dis-	Tot
---	---	---	---
T+	30	8	38
T-	7	22	29
Tot	37	30	67 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	81.0%	68.4%	93.6%
Sp	73.0%	57.1%	88.9%
PPV	78.9%	66.0%	91.9%
NPV	75.9%	60.3%	91.4% | 2) MRI |	Dis+	Dis-	Tot
---	---	---	---
T+	36	5	41
T-	1	25	26
Tot	37	30	67 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	97.3%	92.1%	100.0%
Sp	83.3%	70.0%	96.7%
PPV	87.8%	77.8%	97.8%
NPV	96.2%	88.8%	100.0%		**Comments::** --Unclear how cases selected --Clinical presentation not described --No scoring system for US – unclear if descriptive analysis required all or any of the findings to be considered malignant – MRI more clearly described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	• Wall thickness > 3 mm or nodular or other solid components • Entirely solid • Involvement of adjacent organs • Ascites • Lymph node > 1 cm • Peritoneal, mesenteric, or omental disease				
Menon, Talaat, Rosenthal, et al., 2000 **#2780**	**Geographical location:** London, UK **Dates:** 1986-1989 **Size of population:** 22,000 in prevalence screen, 10,958 randomized to 3 annual incidence screens. Results based on 741women **Screening study** **Reference standard:** Surgery, registry diagnosis of cancer, followup questionnaire **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced or discussed **Statistical tests used:**	**Age:** NR **Menopausal status (n [%]):** 100% postmenopausal, > 45 **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** All with elevated CA-125 **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** All with CA-125 > 30 75.3% of scans transabdominal, 8.4% transvaginal, 16.2% both	All results given using number of scans as denominator (n = 1219), disease defined as incident cancer within 1 year of scan 1) Abnormal volume as threshold Dis+ / Dis- / Tot T+: 17 / 63 / 80 T-: 2 / 945 / 947 Tot: 19 / 1008 / 1027 Value (95% CI): Se 89.5% (75.7%–100.0%); Sp 93.8% (92.3%–95.2%); PPV 21.3% (12.3%–30.2%); NPV 99.8% (99.5%–100.0%) 2) Abnormal morphology as threshold Dis+ / Dis- / Tot T+: 19 / 61 / 80 T-: 0 / 947 / 947 Tot: 19 / 1008 / 1027 Value (95% CI): Se 100.0% (84.2%–100.0%); Sp 93.9% (92.5%–95.4%); PPV 23.8% (14.4%–33.1%); NPV 100.0% (99.7%–100.0%)	**Comments:** --Authors state no significant differences in sensitivity, but underpowered to detect differences **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - (large study, but small number of cancers) Statistical tests: - Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Se/Sp **Blinding:** Yes **Definition of positive and negative on screening test:** If CA-125 > 30 U/ml, ultrasound If ultrasound normal: < 8.8 ml volume, uniform hypoechogenicity, smooth outlines, or not visualized but no abnormality repeated CA-125 q 3 months x 1 year, then annual screening: Equivocal: Ovarian volume < 8.8 ml with abnormal morphology: repeat scan q 6 weeks until either normal or abnormal scan Abnormal: Ovarian volume > 8.8 ml, referred to gynecologist, with management at their discretion			3) Complex morphology as threshold Dis+ Dis- Tot T+ **16** **27** 43 T- **3** **981** 984 Tot 19 1008 1027 Value / Lower 95% CI / Upper 95% CI Se 84.2% 67.8% 100.0% Sp 97.3% 96.3% 98.3% PPV 37.2% 22.8% 51.7% NPV 99.7% 99.4% 100.0%	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Merce, Caballero, Barco, et al., 1998 **#3510**	**Geographical location:** Spain **Dates:** 1990 - 1995 **Size of population:** 213 129 had surgery **Other** Patients with masses ?referred to hospital **Reference standard:** Histopathology for those in surgery, repeat US for the others **Reference standard applied to all test negatives?:** No **Test reliability established?:** Unclear for US score Yes for RI **Statistical tests used:** ROC curves **Blinding:** NR but US prior to surgery **Definition of positive and negative on screening test:** RI ≤ 0.5 For US sonographic index score ≥ 6 which was volume + morphology score	**Age:** Mean: 41 Range: 15 - 87 **Menopausal status** (n [%]): NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Premenopausal women with ovarian mass >28mm in diameter, or >10mm for menopausal women. **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US score ≥ 6 Dis+ / Dis- / Tot T+: 21 / 34 / 55 T-: 1 / 73 / 74 Tot: 22 / 107 / 129 Value (Lower 95% CI – Upper 95% CI) Se: 95.5% (86.8% – 100.0%) Sp: 68.2% (59.4% – 77.0%) PPV: 38.2% (25.3% – 51.0%) NPV: 98.6% (96.0% – 100.0%) 2) RI ≤ 0.5 Dis+ / Dis- / Tot T+: 18 / 36 / 54 T-: 4 / 71 / 75 Tot: 22 / 107 / 129 Value (Lower 95% CI – Upper 95% CI) Se: 80.0% (63.3% – 96.7%) Sp: 66.7% (57.8% – 75.6%) PPV: 33.3% (20.8% – 45.9%) NPV: 94.7% (89.6% – 99.8%)	**Comments:** --Time interval from original diagnosis of mass to followup US was 2 weeks to 3 months --US score developed by authors in different paper [ref 15] and unclear how reliable. Test features of score not described. --2x2 tables based on the 129 who had surgery only **Quality assessment:** Reference standard: +, 2x2 table based only on those who had surgery Verification bias: -, only US followup in 213-129 patients Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Morgante, la Marca, Ditto, et al., 1999 **#2900**	**Geographical location:** Siena, Italy **Dates:** Jan 1995 – Dec 1997 **Size of population:** 124 **Other** Case series Consecutive admissions for surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp, ROC Chi-square Mann-Whitney U test **Blinding:** NR - prospective **Definition of positive and negative on screening test:** CA-125 – range used US score: Presence of multilocular cystic lesions (1), solid areas (1), bilateral lesions (1), ascites (1), intraabdominal mets (1)	**Age:** NR: **Menopausal status (n [%]):** Pre (< 45): 69 (55.6%) Post (> 55): 55 (44.3%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age >30 Mass scheduled for surgery in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US score of 2 |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 22 | 12 | 34 | | T- | 9 | 81 | 90 | | Tot | 31 | 93 | 124 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 71.0% | 55.0% | 87.0% | | Sp | 87.0% | 80.2% | 93.8% | | PPV | 64.7% | 48.6% | 80.8% | | NPV | 90.0% | 83.8% | 96.2% | 2) CA-125 ≥ 25 U/ml |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 24 | 15 | 39 | | T- | 7 | 78 | 85 | | Tot | 31 | 93 | 124 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 78.0% | 63.4% | 92.6% | | Sp | 84.0% | 76.5% | 91.5% | | PPV | 61.5% | 46.3% | 76.8% | | NPV | 91.8% | 85.9% | 97.6% | 3) CA-125 ≥ 50 U/ml |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 23 | 5 | 28 | | T- | 8 | 88 | 96 | | Tot | 31 | 93 | 124 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 74.0% | 58.6% | 89.4% | | Sp | 95.0% | 90.6% | 99.4% | | PPV | 82.1% | 68.0% | 96.3% | | NPV | 91.7% | 86.1% | 97.2% |	**Comments:** --What constitutes + CA-125 or US score not discussed --Borderline tumors grouped in with malignant in analysis **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: +/- Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
O'Connell, Ryan, Murphy, et al., 1987 **#6690**	**Geographical location:** Hamilton, Ontario **Dates:** Nov 1984-May 1986 **Size of population:** 56 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se/Sp/ROC **Blinding:** Yes **Definition of positive and negative on screening test:** Varied	**Age:** Mean (SD): 55 Range: 13-81 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** "High clinical suspicion of ovarian cancer" CA-125 drawn within 3 days of surgery	1) CA-125 threshold >35 U/ml, ovarian cancer vs. any other diagnosis (including other malignancy) Dis+ / Dis- / Tot T+: 26 / 17 / 43 T-: 0 / 13 / 13 Tot: 26 / 30 / 56 Value / Lower 95% CI / Upper 95% CI Se: 100.0% / 88.5% / 100.0% Sp: 43.3% / 25.6% / 61.1% PPV: 60.5% / 45.9% / 75.1% NPV: 100.0% / 76.9% / 100.0% 2) CA-125 threshold > 60 U/ml, ovarian cancer vs. any other diagnosis (including other malignancy) Dis+ / Dis- / Tot T+: 24 / 12 / 36 T-: 2 / 18 / 20 Tot: 26 / 30 / 56 Value / Lower 95% CI / Upper 95% CI Se: 92.0% / 81.6% / 100.0% Sp: 60.0% / 42.5% / 77.5% PPV: 66.7% / 51.3% / 82.1% NPV: 90.0% / 76.9% / 100.0% 3) CA-125 threshold > 35 U/ml, any malignancy vs. benign Dis+ / Dis- / Tot T+: 36 / 7 / 43 T-: 2 / 11 / 13 Tot: 38 / 18 / 56 Value / Lower 95% CI / Upper 95% CI Se: 95.0% / 88.1% / 100.0% Sp: 61.0% / 38.5% / 83.5%	**Comments:** --Clinical presentation not reported --Very high prevalence of cancer --Borderline masses included in malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: - Blinding: + Definition of +/- on screening test: + **Comments:** --Clinical presentation not reported --Very high prevalence of cancer --Borderline masses included in malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: - Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

PPV	83.7%	72.7%	94.8%
NPV	84.6%	65.0%	100.0%

4) CA-125 threshold > 60 U/ml, any malignancy vs. benign

	Dis+	Dis-	Tot
T+	33	3	36
T-	5	15	20
Tot	**38**	**18**	56

	Value	Lower 95% CI	Upper 95% CI
Se	**87.0%**	76.3%	97.7%
Sp	**83.0%**	65.6%	100.0%
PPV	91.7%	82.6%	100.0%
NPV	75.0%	56.0%	94.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Onsrud, Shabana, and Austgulen, 1996 **#3950**	**Geographical location:** Trondheim, Norway **Dates:** NR **Size of population:** 72 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Wilcoxon rank sum, Se/Sp **Blinding:** Yes **Definition of positive and negative on screening test:** Varied	**Age:** Range: 19-80 years **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Stage I: 20/45 (44.4%) Stage II: 1/45 (2.2%) Stage III: 21/45 (46.7%) Stage IV: 1/45 (2.2%)	1) CA-125 > 20 U/ml Dis+ / Dis- / Tot T+: 37 / 4 / 41 T-: 8 / 23 / 31 Tot: 45 / 27 / 72 Value (95% CI) Se: 82.0% (70.8% - 93.2%) Sp: 85.0% (71.5% - 98.5%) PPV: 90.2% (81.2% - 99.3%) NPV: 74.2% (58.8% - 89.6%) 2) TNF p 55 > 2.0 ng/ml Dis+ / Dis- / Tot T+: 26 / 3 / 29 T-: 19 / 24 / 43 Tot: 45 / 27 / 72 Value (95% CI) Se: 58.0% (43.6% - 72.4%) Sp: 89.0% (77.2% - 100.0%) PPV: 89.7% (78.6% - 100.0%) NPV: 55.8% (41.0% - 70.7%) 3) TNF p 75 > 4.3 ng/ml Dis+ / Dis- / Tot T+: 7 / 1 / 8 T-: 38 / 26 / 64 Tot: 45 / 27 / 72 Value (95% CI) Se: 16.0% (5.3% - 26.7%) Sp: 96.0% (88.6% - 100.0%) PPV: 87.5% (64.6% - 100.0%) NPV: 40.6% (28.6% - 52.7%)	**Comments:** --Clinical presentation not reported --Spectrum of disease reported --Patient selection criteria not described -- CA-125 cutpoint derived from mean +2 SD of the 26 control women **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: + **Quality Scoring**

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Padungsutt, Thira-pagawong, Senapad, et al., 2000 **#520**	**Geographical location:** Bangkok, Thailand **Dates:** May 1996-Mar 1997 **Size of population:** 92 women **Other** Case series **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** **Blinding:** **Definition of positive and negative on screening test:**	**Age:** Mean (SD): 38.8 (2.8) Range: 10-69 **Menopausal status (n [%]):** Pre (< 45): 64 (69.6%) Post (> 55): 28 (30.4%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Stage I: 17/40 (42.5%) Stage II: 3/40 (7.5%) Stage III: 20/40 (50%)	1) Tissue polypeptide specific antigen > 80 U/L Dis+ / Dis- / Tot T+: 36 / 15 / 51 T-: 4 / 37 / 41 Tot: 40 / 52 / 92 Value / Lower 95% CI / Upper 95% CI Se: 90.0% / 80.7% / 99.3% Sp: 71.2% / 58.8% / 83.5% PPV: 70.6% / 58.1% / 83.1% NPV: 90.2% / 81.2% / 99.3%	**Comments:** --Clinical presentation not described --Not stratified by age, menopausal status --Patient selection criteria not described --Borderline tumors grouped with malignant --Not a common test modality **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:+ Sample size: - Statistical tests:+ Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Patsner and Mann, 1988 **#5360**	**Geographical location:** Stony Brook, NY **Dates:** Jul 1985-Jul 1987 **Size of population:** 250 women **Other** Consecutive case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se/Sp **Blinding:** Yes **Definition of positive and negative on screening test:** CA-125 > 35	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 125 (50%) Post (> 55): 125 (50%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Invasive ovarian cancer Stage I: 14/80 (17.5%) Stage II: 3/80 (3.8%) Stage III: 50/80 (62.5%) Stage IV: 13/80 (16.3%)	1) CA-125 > 35 U/ml, All patients |	Dis+	Dis-	Tot
---	---	---	---
T+	92	25	117
T-	36	97	133
Tot	128	122	250 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	71.9%	64.1%	79.7%
Sp	79.5%	72.3%	86.7%
PPV	78.6%	71.2%	86.1%
NPV	72.9%	65.4%	80.5% | 2) CA-125 > 35 U/ml, Premenopausal |	Dis+	Dis-	Tot
---	---	---	---
T+	33	16	49
T-	18	58	76
Tot	51	74	125 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	65.0%	51.9%	78.1%
Sp	78.0%	68.6%	87.4%
PPV	67.3%	54.2%	80.5%
NPV	76.3%	66.8%	85.9% | 3) CA-125 > 35 U/ml, Postmenopausal |	Dis+	Dis-	Tot
---	---	---	---
T+	59	9	68
T-	18	39	57
Tot	77	48	125 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	77.0%	67.6%	86.4%
Sp	81.0%	69.9%	92.1%
PPV	86.8%	78.7%	94.8%
NPV	68.4%	56.4%	80.5%	**Comments:** --Clinical presentation not described --Borderline tumors grouped with malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Prompeler, Madjar, and Sauerbrei, 1996 **#3960**	**Geographical location:** Freiburg, Germany **Dates:** Jul 1992-Jul 1994 **Size of population:** 212 consecutive cases **Other** Case series **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Discussed **Statistical tests used:** Spearman's, Mann-Whitney, See/Sp **Blinding:** No **Definition of positive and negative on screening test:** 20 different measures of ovarian vascularity, flow studied. 4 best criteria were • Minimum resistive index (difference of peak systolic and maximum end diastolic, divided by peak systolic flow velocity; cutpoint 0.5	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 81 (38.2%) Post (> 55): 131 (61.8%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** US timed to surgery date – therefore not all premenopausal US done in proliferative phase	Malignant disease includes borderline (n = 9) 1) Total number of arteries > 4, all patients Dis+ / Dis- / Tot T+: 56 / 34 / 90 T-: 12 / 110 / 122 Tot: 68 / 144 / 212 Value (95% CI): Se 82.4% (73.3%-91.5%); Sp 76.4% (69.5%-83.3%); PPV 62.2% (52.2%-72.2%); NPV 90.2% (84.9%-95.4%) 2) Total number of arteries > 4, postmenopausal Dis+ / Dis- / Tot T+: 45 / 14 / 59 T-: 10 / 62 / 72 Tot: 55 / 76 / 131 Value (95% CI): Se 81.8% (71.6%-92.0%); Sp 81.6% (72.9%-90.3%); PPV 76.3% (65.4%-87.1%); NPV 86.1% (78.1%-94.1%) 3) Total number of arteries > 4, premenopausal Dis+ / Dis- / Tot T+: 11 / 20 / 31 T-: 2 / 48 / 50 Tot: 13 / 68 / 81 Value (95% CI): Se 84.6% (65.0%-100.0%); Sp 70.6% (59.8%-81.4%)	**Comments:** --Clinical presentation not described --LMP tumors grouped with malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: + (discussion of power) Statistical tests: + Blinding: Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	• Total number of tumor arteries in color mode; cutpoint 4 • Maximum systolic flow velocity; cutpoint 30 cm/sec • Sum of all systolic flow velocities; cutpoint 75 cm/sec			PPV 35.5% 18.6% 52.3% NPV 96.0% 90.6% 100.0% 4) Min resistive index > 0.5, all patients Dis+ Dis- Tot T+ 56 59 115 T- 12 85 97 Tot 68 144 212 Value Lower 95% CI Upper 95% CI Se 82.1% 73.0% 91.2% Sp 59.2% 51.2% 67.2% PPV 48.7% 39.6% 57.8% NPV 87.6% 81.1% 94.2% 5) Min resistive index, postmenopausal Dis+ Dis- Tot T+ 45 23 68 T- 10 53 63 Tot 55 76 131 Value Lower 95% CI Upper 95% CI Se 81.5% 71.2% 91.8% Sp 69.2% 58.8% 79.6% PPV 66.2% 54.9% 77.4% NPV 84.1% 75.1% 93.2% 6) Min resistive index, premenopausal Dis+ Dis- Tot T+ 11 36 47 T- 2 32 34 Tot 13 68 81 Value Lower 95% CI Upper 95% CI Se 84.6% 65.0% 100.0% Sp 47.1% 35.2% 58.9% PPV 23.4% 11.3% 35.5% NPV 94.1% 86.2% 100.0%	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

7) Maximum systolic velocity, all patients

	Dis+	Dis-	Tot
T+	54	33	87
T-	14	111	125
Tot	68	144	212

	Value	Lower 95% CI	Upper 95% CI
Se	79.1%	69.4%	88.8%
Sp	76.8%	69.9%	83.7%
PPV	62.1%	51.9%	72.3%
NPV	88.8%	83.3%	94.3%

8) Max systolic velocity, postmenopausal

	Dis+	Dis-	Tot
T+	42	9	51
T-	13	67	80
Tot	55	76	131

	Value	Lower 95% CI	Upper 95% CI
Se	75.9%	64.6%	87.2%
Sp	87.7%	80.3%	95.1%
PPV	82.4%	71.9%	92.8%
NPV	83.8%	75.7%	91.8%

9) Max systolic velocity, premenopausal

	Dis+	Dis-	Tot
T+	21	24	45
T-	1	44	45
Tot	22	68	90

	Value	Lower 95% CI	Upper 95% CI
Se	95.5%	86.8%	100.0%
Sp	64.7%	53.3%	76.1%
PPV	46.7%	32.1%	61.2%
NPV	97.8%	93.5%	100.0%

10) Sum of systolic velocities, all patients

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	61	38	99
T-	7	106	113
Tot	68	144	212

	Value	Lower 95% CI	Upper 95% CI
Se	89.6%	82.3%	96.9%
Sp	73.6%	66.4%	80.8%
PPV	61.6%	52.0%	71.2%
NPV	93.8%	89.4%	98.2%

11) Sum of systolic velocities, postmenopausal

	Dis+	Dis-	Tot
T+	50	12	62
T-	5	64	69
Tot	55	76	131

	Value	Lower 95% CI	Upper 95% CI
Se	90.7%	83.0%	98.4%
Sp	84.6%	76.5%	92.7%
PPV	80.6%	70.8%	90.5%
NPV	92.8%	86.6%	98.9%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Pyrgiotis, Sala-malekis, Loghis, et al., 1993 **#4790**	**Geographical location:** Athens, Greece **Dates:** NR **Size of population:** 126 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se/Sp **Blinding:** Yes **Definition of positive and negative on screening test:** CA-125 > 35 U/ml	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) All pelvic masses Borderline included as malignant (n = 4) CA-125 > 35 U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	48	7	55
T-	14	57	71
Tot	62	64	126 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	77.4%	67.0%	87.8%
Sp	89.1%	81.4%	96.7%
PPV	87.3%	78.5%	96.1%
NPV	80.3%	71.0%	89.5% | 2) Ovarian masses only (excluding fibroids etc) |	Dis+	Dis-	Tot
---	---	---	---
T+	46	3	49
T-	8	26	34
Tot	54	29	83 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	85.2%	75.7%	94.7%
Sp	89.7%	78.6%	100.0%
PPV	93.9%	87.2%	100.0%
NPV	76.5%	62.2%	90.7%		**Comments:** --Clinical presentation not described --Drawn within 2 days of surgery **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Rehn, Lohmann, and Rempen, 1996 **#3910**	**Geographical location:** Wurzburg, Germany **Dates:** Mar 1992 – Dec 1994 **Size of population:** 310 **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes for 300 women 10 premenopausal women were "followed up to resolution" **Test reliability established?:** Yes **Statistical tests used:** Chi square U test Wilcoxon test Regression analysis **Blinding:** NR **Definition of positive and negative on screening test:** US morphology – Sassone's criteria score ≥ 9 as well as presence of "inhomogeneous echogenicity, irregular wall, or solid portion" considered suggestive of	**Age:** Mean: 43.5 Range: 17-88 **Menopausal status (n [%]):** Pre (< 45): 227 (73%) Post (> 55): 83(27%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Unclear – presumably women with masses in hospital during time frame who had US **Exclusion criteria:** Uterine malignancy	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 310(100%) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Sassone's score ≥ 9 Dis+ / Dis- / Tot T+: 43 / 70 / 113 T-: 8 / 189 / 197 Tot: 51 / 259 / 310 Value (95% CI) Se: 84.0% (73.9% – 94.1%) Sp: 73.0% (67.6% – 78.4%) PPV: 38.1% (29.1% – 47.0%) NPV: 95.9% (93.2% – 98.7%) 2) US morphology descriptive (group 4) Dis+ / Dis- / Tot T+: 46 / 91 / 137 T-: 5 / 168 / 173 Tot: 51 / 259 / 310 Value (95% CI) Se: 90.0% (81.8% – 98.2%) Sp: 65.0% (59.2% – 70.8%) PPV: 33.6% (25.7% – 41.5%) NPV: 97.1% (94.6% – 99.6%) 3) PI ≤ 1.0 Dis+ / Dis- / Tot T+: 34 / 122 / 156 T-: 17 / 137 / 154 Tot: 51 / 259 / 310 Value (95% CI) Se: 67.0% (54.1% – 79.9%) Sp: 53.0% (46.9% – 59.1%) PPV: 21.8% (15.3% – 28.3%) NPV: 89.0% (84.0% – 93.9%)	**Comments:** --Document overlap in RI between benign and malignant masses --Unclear if prospective or retrospective case series --LMP tumors grouped with malignant --Overlap noted for PI between malignant and benign masses --Regression analysis showed negative correlation of PI and tumor size for malignant masses and in total, but independence when only benign lesions analysed --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	malignancy PI ≤ 1.0				
Reles, Wein, and Lichtenegger, 1997 **#6090**	**Geographical location:** Berlin, Germany **Dates:** Mar 1992-Aug 1994 **Size of population:** 98 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Chi-square, Mann-Whitney, Se/Sp **Blinding:** NR **Definition of positive and negative on screening test:** Morphology: Malignant if • Complex cystic pattern with irregularly thick septae • Cystic or polycystic	**Age:** NR **Menopausal status (n [%]):** Pre (< 45):33 (36.3%) Post (> 55): 52 (57.1%) Unknown: 5 (5.55%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Cancers of non-ovarian origin (n = 7) not included 1) TVUS, all patients |	Dis+	Dis-	Tot
---	---	---	---
T+	20	11	31
T-	2	58	60
Tot	22	69	91 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	90.9%	78.9%	100.0%
Sp	84.1%	75.4%	92.7%
PPV	64.5%	47.7%	81.4%
NPV	96.7%	92.1%	100.0% | 2) TVUS, premenopausal (n = 33) |	Dis+	Dis-	Tot
---	---	---	---
T+	5	6	11
T-	0	22	22
Tot	5	28	33 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	40.0%	100.0%
Sp	78.6%	63.4%	93.8%
PPV	45.5%	16.0%	74.9%
NPV	100.0%	86.4%	100.0% | 3) TVUS, postmenopausal (n = 52) |	Dis+	Dis-	Tot
---	---	---	---
T+	13	4	17
T-	2	33	35
Tot	15	37	52		**Comments:** --Clinical presentation not described -- 2x2 tables for pre and post menopausal subgroups are approximate, with N, Se, Sp consistent with Table 3, but allowing for some discrepancies with PPV or NPV **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests:+ Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

pattern with papillary or indented mural part
- Polycystic pattern with irregularly thick septae and solid part < 50%
- Solid pattern (>50%) with irregular cystic part
- Completely solid homogeneous or inhomogeneous patter

(modified morphology classification from Vera 1986 and Kawai 1992)

Color Doppler
Pulsatility index ≤ 1.1
Resistive index 0.7
Lowest measured PI and RI used

	Value	Lower 95% CI	Upper 95% CI
Se	86.7%	69.5%	100.0%
Sp	89.2%	79.2%	99.2%
PPV	76.5%	56.3%	96.6%
NPV	94.3%	86.6%	100.0%

4) Color Doppler pulsatility index ≤ 1.1

	Dis+	Dis-	Tot
T+	18	12	30
T-	2	34	36
Tot	20	46	66

	Value	Lower 95% CI	Upper 95% CI
Se	90.0%	76.9%	100.0%
Sp	74.0%	61.3%	86.7%
PPV	60.1%	42.5%	77.6%
NPV	94.5%	87.0%	100.0%

5) Color Doppler pulsatility index ≤ 1.1, premenopausal (n = 24)

	Dis+	Dis-	Tot
T+	4	6	10
T-	1	13	14
Tot	5	19	24

	Value	Lower 95% CI	Upper 95% CI
Se	80.0%	44.9%	100.0%
Sp	68.4%	47.5%	89.3%
PPV	40.0%	9.6%	70.4%
NPV	92.9%	79.4%	100.0%

6) Color Doppler pulsatility index ≤ 1.1, postmenopausal (n = 37)

	Dis+	Dis-	Tot
T+	13	4	17
T-	1	19	20
Tot	14	23	37

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				Value / Lower 95% CI / Upper 95% CI Se 92.9% 79.4% 100.0% Sp 82.6% 67.1% 98.1% PPV 76.5% 56.3% 96.6% NPV 95.0% 85.4% 100.0%	
Reuter, Steffens, Schuppler, et al., 1998 **#10990**	**Geographical location:** Germany **Dates:** Jan 1994 – Aug 1995 **Size of population:** 65 **Case series (retrospective)** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Chi square **Blinding:** NR **Definition of positive and negative on screening test:** Used criteria from reference [16] Malignancy assumed if	**Age:** Mean: 48.8 Range: 18-84 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients with suspected adnexal tumors in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) MRI Dis+ / Dis- / Tot T+ 12 12 24 T- 0 41 41 Tot 12 53 65 Value / Lower 95% CI / Upper 95% CI Se 100.0% 75.0% 100.0% Sp 78.2% 67.1% 89.3% PPV 50.0% 30.0% 70.0% NPV 100.0% 92.7% 100.0% 2) TVUS Dis+ / Dis- / Tot T+ 12 18 30 T- 0 35 35 Tot 12 53 65 Value / Lower 95% CI / Upper 95% CI Se 100.0% 75.0% 100.0% Sp 65.5% 52.7% 78.3% PPV 40.0% 22.5% 57.5% NPV 100.0% 91.4% 100.0%	**Comments:** --For MRI used different machines, different contrast materials, different imaging techniques with study group --Diagnosis done by consensus in conference (2 gynecologists for US, 2 radiologists for MRI) – not blinded to each other --No mention of whether blinded to pathologic results --Used criteria for malignancy (citation 16) which are not common --Borderline grouped with malignant -TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: +/- Blinding: - Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	one or more of the following was present: solid growth or solid component wall with or without necrosis; cystic lesion with thickness of walls or septa more than 3 mm; nodular or papillary projections; excessive multilocularity; infiltration in neighboring organs or pelvic wall; tumor manifestation in the peritoneum, mesentery, omentum; lymphadenopathy. Benign was none of the above present				
Roman, Muder-spach, Burnett, et al., 1998 **#3410**	**Geographical location:** Los Angeles, CA **Dates:** Jul 1992-Mar 1994 **Size of population:** 226 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Fisher's exact test, Se/Sp	**Age:** Mean: 39 Range: 13-80 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Isolated pelvic mass with surgery planned **Exclusion criteria:** Emergency surgery Evidence of metastases Past history of cancer	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CEA > 3 ng/ml, borderline classified as malignant Dis+ / Dis- / Tot T+: **7** / **13** / 20 T-: **25** / **181** / 206 Tot: 32 / 194 / 226 Value / Lower 95% CI / Upper 95% CI Se: 21.9% / 7.6% / 36.2% Sp: 93.3% / 89.8% / 96.8% PPV: 35.0% / 14.1% / 55.9% NPV: 87.9% / 83.4% / 92.3% 2) CEA > 3 ng/ml, borderline classified as benign Dis+ / Dis- / Tot T+: **5** / **15** / 20 T-: **17** / **189** / 206 Tot: 22 / 204 / 226 Value / Lower 95% CI / Upper 95% CI	**Comments:** --Clinical presentation not described --Not stratified by age or menopausal status --Borderline tumors included with malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				Se 22.7% 5.2% 40.2% Sp 92.6% 89.1% 96.2% PPV 25.0% 6.0% 44.0% NPV 91.7% 88.0% 95.5%	
	Blinding: Yes **Definition of positive and negative on screening test:** CEA: > 3.0 ng/ml in nonsmokers, > 5.0 ng/ml in smokers				
Roman, Muder-spach, Stein, et al., 1997 **#6160**	**Geographical location:** Los Angeles, CA **Dates:** Jul 1992-Mar 1994 **Size of population:** 226 **Other** Non-consecutive case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** **Statistical tests used:** Chi-square, logistic regression, Se/Sp **Blinding:** NR **Definition of positive and negative on**	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 181 (80.1%) Post (> 55): 45 (19.9%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Isolated pelvic mass with surgery planned **Exclusion criteria:** Emergency surgery Evidence of metastases Past history of cancer	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Borderline classified as malignant CA-125 > 35 U/ml Dis+ Dis- Tot T+ 29 54 83 T- 14 129 143 Tot 43 183 226 Value Lower 95% CI Upper 95% CI Se 67.4% 53.4% 81.4% Sp 70.5% 63.9% 77.1% PPV 34.9% 24.7% 45.2% NPV 90.2% 85.3% 95.1% 2) Any positive tumor marker Dis+ Dis- Tot T+ 31 55 86 T- 12 128 140 Tot 43 183 226 Value Lower 95% CI Upper 95% CI Se 72.1% 58.7% 85.5% Sp 70.0% 63.4% 76.6% PPV 36.1% 25.9% 46.2% NPV 91.4% 86.8% 96.1% 3) Any positive tumor marker, post-menopausal	**Comments:** --Clinical presentation not described --US authors stated scoring system used, however, description in text makes it seem more of a descriptive not numeric score **Quality assessment:** Reference standard:+ Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

screening test:
CA-125: >35 U/mL
HCG: > 15
AFP: > 10 ng/ml
LDH: > 350 U/L

Gray scale ultrasound:
Cystic with one large (> 2 cm) or multiple nodules or cystic/solid; completely solid masses not appearing to arise from the uterus in postmenopausal women
Simplified scoring system used (not described in scoring fashion)

Doppler (only if gray scale suspicious)
PI < 1.0 or
RI < 0.4

	Dis+	Dis-	Tot
T+	11	6	17
T-	5	23	28
Tot	16	29	45

	Value	Lower 95% CI	Upper 95% CI
Se	68.8%	46.1%	91.5%
Sp	79.3%	64.6%	94.0%
PPV	64.7%	42.0%	87.4%
NPV	82.1%	68.0%	96.3%

4) Ultrasound—all patients

	Dis+	Dis-	Tot
T+	38	24	62
T-	5	159	164
Tot	43	183	226

	Value	Lower 95% CI	Upper 95% CI
Se	88.4%	78.8%	98.0%
Sp	86.9%	82.0%	91.8%
PPV	61.3%	49.2%	73.4%
NPV	97.0%	94.3%	99.6%

5) Ultrasound: Premenopausal

	Dis+	Dis-	Tot
T+	25	13	38
T-	2	141	143
Tot	27	154	181

	Value	Lower 95% CI	Upper 95% CI
Se	92.6%	82.7%	100.0%
Sp	91.6%	87.2%	96.0%
PPV	65.9%	50.8%	81.0%
NPV	98.6%	96.7%	100.0%

6) Ultrasound: Postmenopausal

	Dis+	Dis-	Tot
T+	13	11	24
T-	3	18	21

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Tot 16 29 45

	Value	Lower 95% CI	Upper 95% CI
Se	81.3%	62.2%	100.0%
Sp	62.1%	44.4%	79.8%
PPV	54.2%	34.3%	74.1%
NPV	85.8%	70.8%	100.0%

Salem, White, and Lai, 1994

#4430

Geographical location:
Toronto, Canada

Dates: Sep 1992 – Jan 1994

Size of population:
99 patients
102 masses from among 377 women with adnexal mass at sonography

Other
Referral series

Reference standard:
Histopathology

Reference standard applied to all test negatives?:
No

Test reliability established?:
US – probably Sassone PI yes

Statistical tests used:
Se, Spec

Blinding:
Yes

Age:
Range: 15-58

Menopausal status (n [%]):
Pre (< 45): 57 (57.6%)
Peri (45-54): 23 (23.2%)
Post (> 55): 19 (19.2%)

Race/ethnicity (n [%]):
NR

Risk factors (n [%]):
NR

Inclusion criteria:
Women with diagnosis of mass referred for surgery

Exclusion criteria:
Studies done not in first 10 days after menstruation in premenopausal women (to avoid luteal phase)
PI that couldn't be measured due to no flow (present in 7 masses)

Symptomatic (n [%]):
NR

Detected by exam (n [%]):
NR

Detected by imaging (n [%]):
100% had adnexal mass detected at sonography

Combination (n [%]):
NR

Additional data used for diagnosis:
NR

1) PI < 1

	Dis+	Dis-	Tot
T+	10	17	27
T-	3	65	68
Tot	13	82	95

	Value	Lower 95% CI	Upper 95% CI
Se	76.9%	54.0%	99.8%
Sp	79.3%	70.5%	88.0%
PPV	37.0%	18.8%	55.3%
NPV	95.6%	90.7%	100.0%

2) PI < 1, peri- and postmenopausal patients

	Dis+	Dis-	Tot
T+	8	9	17
T-	3	22	25
Tot	11	31	42

	Value	Lower 95% CI	Upper 95% CI
Se	72.7%	46.4%	99.0%
Sp	71.0%	55.0%	86.9%
PPV	47.1%	23.3%	70.8%
NPV	88.0%	75.3%	100.0%

3) PI < 1, premenopausal patients

	Dis+	Dis-	Tot
T+	2	8	10
T-	0	43	43

Comments:
--Analysis done on masses not individuals
--No flow detected in 7 of 89 benign lesions; these cases excluded.
--Scoring system most likely Sasonne or modified but not described at all.
--But unable to use US morphology to construct 2x2 tables

Quality assessment:
Reference standard: +
Verification bias: -, only 99 women had histopathol of 377 women with adnexal masses at US
Test reliability/variability: +
Sample size: -
Statistical tests: +
Blinding: +
Definition of +/- on screening test: + for PI

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Definition of positive and negative on screening test:
Pulsatility index (PI) = (PSV-EDV)/mean velocity. Considered positive if < 1.

Tot	2	51	53

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	0.0%	100.0%
Sp	84.3%	74.3%	94.3%
PPV	20.0%	0.0%	44.8%
NPV	100.0%	93.0%	100.0%

4) PI < 1, age > 45 years

	Dis+	Dis-	Tot
T+	**8**	**3**	11
T-	**9**	**22**	31
Tot	17	25	42

	Value	Lower 95% CI	Upper 95% CI
Se	47.1%	23.3%	70.8%
Sp	88.0%	75.3%	100.0%
PPV	72.7%	46.4%	99.0%
NPV	71.0%	55.0%	86.9%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Sassone, Timor-Tritsch, Artner, et al., 1991 **#6780**	**Geographical location:** New York, NY **Dates:** Jun 1987 – Dec 1989 **Size of population:** 143 women **Retrospective case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Se, Sp Correlation coefficient for US and pathology **Blinding:** Yes **Definition of positive and negative on screening test:** Sassone ≥ 9	**Age:** Mean: 41.6 Median: 41 Range: 20-85 **Menopausal status (n [%]):** Pre (< 45): 116(80.4%) Post (> 55): 24(16.8%) 3 unknown secondary to hysterectomy (2.1%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** All laparotomy performed for gynecologic indications in hospital in time frame **Exclusion criteria:** Pregnancy Previous BSO Previous treated carcinoma	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Sassone ≥ 9 |	Dis+	Dis-	Tot
---	---	---	---
T+	13	22	35
T-	0	108	108
Tot	13	130	143 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	76.9%	100.0%
Sp	83.0%	76.5%	89.5%
PPV	37.1%	21.1%	53.2%
NPV	100.0%	97.2%	100.0%		**Comments:** --This is the original article describing the Sassone scoring system – good description of criteria with photos --Borderline tumors grouped in with benign --No mention of clinical pathway of patients --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: +/- Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Sawicki, Spiewan-kiewicz, Cendrowski, et al., 2001 **#2450**	**Geographical location:** Warsaw, Poland **Dates:** NR **Size of population:** 329 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Chi-square, Fishers's exact test, Mann-Whitney, t-test, Se/Sp **Blinding:** Not described **Definition of positive and negative on screening test:** Gray-scale ultra sound: (Sassone's criteria) • Solid or mixed solid/cystic • Faded borders • Septae > 3 mm • Solid papillary projections into cyst	**Age:** Range: 15-88 Mean for patients with benign lesions: 42.6 (12.3) Mean for patients with malignant lesions: 53.1 (12.6) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Premenopausal US on day 7-8 Stage I: 20/74 (27%) Stage II: 9/74 (12.2%) Stage III: 42/74 (56.7%) Stage IV: 3/74 (4.1%)	1) TVUS, morphology only Dis+ / Dis- / Tot T+: 64 / 83 / 147 T-: 10 / 172 / 182 Tot: 74 / 255 / 329 Value (Lower 95% CI – Upper 95% CI) Se: 86.5% (78.7% – 94.3%) Sp: 67.5% (61.8% – 73.2%) PPV: 43.6% (35.6% – 51.6%) NPV: 94.5% (91.2% – 97.8%) 2) Color Doppler (unclear if test characteristics calculated based on Doppler alone, or Doppler plus morphology) Dis+ / Dis- / Tot T+: 69 / 14 / 83 T-: 5 / 241 / 246 Tot: 74 / 255 / 329 Value (Lower 95% CI – Upper 95% CI) Se: 93.2% (87.5% – 98.9%) Sp: 94.5% (91.7% – 97.3%) PPV: 83.1% (75.0% – 91.2%) NPV: 98.0% (96.2% – 99.7%)	**Comments:** --Clinical presentation not described --Unclear how Doppler is being compared to morphology—independently or as adjunct --Unclear what was included in "Doppler" measurement – RI, vascularization, and "a subjective semiquantitative assessment of the amount of blood flow (area and color scale)" within lesion **Quality assessment:** Reference standard: + Verification bias:+ Test reliability/variability: - Sample size: - (formal statistical testing of differences in test characteristics, but no discussion of Statistical tests: - (basis of comparison not described) Blinding: Definition of +/- on screening test: - (resistive index threshold given, but not clear how other parameters used)

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	cavity > 3 mm • > 50 mm free fluid in cul de sac Color Doppler: RI < 0.5 Other criteria included presence of "intense" angiogenesis within septae or solid parts of tumor, but unclear how this was used in determining benign vs malignant				
Schelling, Braun, Kuhn, et al., 2000 **#2770**	**Geographical location:** Munich, Germany **Dates:** NR **Size of population:** 63 in development set; 257 in validation set **Other** Development and validation **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Discussed **Statistical tests used:** Logistic regression for development; Se/Sp for validation	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 166 (64.6%) Post (> 55): 91 (35.4%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** History of malignancy	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Ultrasound (any solid component = positive), all patients |	Dis+	Dis-	Tot
---	---	---	---
T+	38	42	80
T-	1	176	177
Tot	39	218	257 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	97.4%	92.5%	100.0%
Sp	80.7%	75.5%	86.0%
PPV	47.5%	36.6%	58.4%
NPV	99.4%	98.3%	100.0% | 2) Ultrasound, any solid component, premenopausal |	Dis+	Dis-	Tot
---	---	---	---
T+	10	25	35
T-	1	130	131
Tot	11	155	166 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	90.9%	73.9%	100.0%
Sp	83.9%	78.1%	89.7%
PPV	28.6%	13.6%	43.5%
NPV	99.2%	97.7%	100.0%		**Comments:** --Clinical presentation not described --Criteria for selecting cases not described --For model development – with N = 65 can you have 12 variables in predictive model and still be stable/valid? Too many variables for a predictive model (including also the Doppler variables(9 additional variables)? **Quality assessment:** Reference standard: + Verification bias:+ Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Blinding:
Yes

Definition of positive and negative on screening test:
Solid component on morphology, presence of central vascularization on Doppler

3) Utrasound, any solid component, postmenopausal

	Dis+	Dis-	Tot
T+	28	17	45
T-	0	46	46
Tot	28	63	91

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	89.3%	100.0%
Sp	73.0%	62.1%	84.0%
PPV	62.2%	48.1%	76.4%
NPV	100.0%	93.5%	100.0%

4) Central vascularity on Doppler in solid component, all patients

	Dis+	Dis-	Tot
T+	36	14	50
T-	3	204	207
Tot	39	218	257

	Value	Lower 95% CI	Upper 95% CI
Se	92.3%	83.9%	100.0%
Sp	93.6%	90.3%	96.8%
PPV	72.0%	59.6%	84.4%
NPV	98.6%	96.9%	100.0%

5) Central vascularity on Doppler in solid component, premenopausal

	Dis+	Dis-	Tot
T+	10	9	19
T-	1	146	147
Tot	11	155	166

	Value	Lower 95% CI	Upper 95% CI
Se	90.9%	73.9%	100.0%
Sp	94.2%	90.5%	97.9%
PPV	52.6%	30.2%	75.1%
NPV	99.3%	98.0%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				6) Central vascularity on Doppler in solid component, postmenopausal Dis+ Dis- Tot T+ 26 5 31 T- 2 58 60 Tot 28 63 91 Value / Lower 95% CI / Upper 95% CI Se 92.9% 83.3% 100.0% Sp 92.1% 85.4% 98.7% PPV 83.9% 70.9% 96.8% NPV 96.7% 92.1% 100.0%	
Schneider, Schneider, Reed, et al., 1993 **#4830**	**Geographical location:** Arizona, USA Academic Center **Dates:** NR **Size of population:** 55 women **Other** "Cross-sectional" referral **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes Sasonne ? Granberg RI - yes **Statistical tests used:**	**Age:** Mean: 53 Median: 53 Range: 10-79 **Menopausal status (n [%]):** Pre: 22 (40%) Post: 33 (60%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women referred to OB/Gyn department with diagnosis of mass and scheduled for surgery already **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** All patients had ultrasound finding of adnexal mass **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI ≤ 0.8 Dis+ Dis- Tot T+ 15 17 32 T- 1 22 23 Tot 16 39 55 Value / Lower 95% CI / Upper 95% CI Se 93.8% 81.9% 100.0% Sp 56.4% 40.8% 72.0% PPV 46.9% 29.6% 64.2% NPV 95.7% 87.3% 100.0% 2) Granberg et al method US Dis+ Dis- Tot T+ 14 6 20 T- 2 33 35 Tot 16 39 55 Value / Lower 95% CI / Upper 95% CI Se 87.5% 71.3% 100.0% Sp 84.6% 73.3% 95.9% PPV 70.0% 49.9% 90.1% NPV 94.3% 86.6% 100.0%	**Comments:** --RI cutoff calculated from this data (not prospective) --RI ≤ 0.8 not uniform in literature (1.0) **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + for CA-125 and Sasonne ? for other US Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Se, Sp **Blinding:** Yes **Definition of positive and negative on screening test:** Sasonne's score Granberg [5] method of 2D US RI ≤ 0.8 CA-125 > 35 U/ml			3) Sasonne (cut off 9) Dis+ / Dis- / Tot T+: 14 / 10 / 24 T-: 2 / 29 / 31 Tot: **16** / **39** / 55 Value (Lower 95% CI – Upper 95% CI) Se: **87.5%** (71.3% – 100.0%) Sp: **74.4%** (60.7% – 88.1%) PPV: 58.3% (38.6% – 78.1%) NPV: 93.5% (84.9% – 100.0%) 4) CA-125 > 35 U/ml Dis+ / Dis- / Tot T+: 12 / 6 / 18 T-: 4 / 33 / 37 Tot: **16** / **39** / 55 Value (Lower 95% CI – Upper 95% CI) Se: **75.0%** (53.8% – 96.2%) Sp: **84.6%** (73.3% – 95.9%) PPV: 66.7% (44.9% – 88.4%) NPV: 89.2% (79.2% – 99.2%)	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Schutter, Davelaar, van Kamp, et al., 2002 **#2160**	**Geographical location:** Amsterdam and Enschede, The Netherlands **Dates:** Nov 1990-Dec 1992 **Size of population:** 511; serum available for 412 **Other** Serum bank **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se/Sp; logistic regression **Blinding:** Yes **Definition of positive and negative on screening test:** Varied; "optimal" cutpoints CA-125: 57 u/mL CA-15-3: 26 u/mL CA-724: 3. 5 u/mL	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125, ovarian cancer vs benign ovarian mass, threshold = 35 u/mL Dis+ / Dis- / Tot T+: 108 / 35 / 143 T-: 25 / 94 / 119 Tot: 133 / 129 / 262 Value (Lower 95% CI – Upper 95% CI) Se: 81.0% (74.3% – 87.7%) Sp: 73.0% (65.3% – 80.7%) PPV: 75.6% (68.5% – 82.6%) NPV: 78.8% (71.5% – 86.2%) 2) CA-125, ovarian cancer vs benign ovarian mass, threshold = 57 u/mL Dis+ / Dis- / Tot T+: 102 / 23 / 126 T-: 31 / 106 / 136 Tot: 133 / 129 / 262 Value (Lower 95% CI – Upper 95% CI) Se: 77.0% (69.8% – 84.2%) Sp: 82.0% (75.4% – 88.6%) PPV: 81.5% (74.7% – 88.3%) NPV: 77.6% (70.6% – 84.6%) 3) CA-15-3, ovarian cancer vs benign mass, threshold = 26 u/mL Dis+ / Dis- / Tot T+: 82 / 9 / 91 T-: 51 / 120 / 171 Tot: 133 / 129 / 262 Value (Lower 95% CI – Upper 95% CI) Se: 62.0% (53.8% – 70.2%) Sp: 93.0% (88.6% – 97.4%) PPV: 90.1% (84.0% – 96.2%) NPV: 70.4% (63.5% – 77.2%)	**Comments:** --Clinical presentation not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:+ Sample size: - Statistical tests: Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

4) CA-724, Ovarian cancer vs benign ovarian mass, threshold = 3.5 u/mL

	Dis+	Dis-	Tot
T+	84	18	102
T-	49	111	160
Tot	133	129	262

	Value	Lower 95% CI	Upper 95% CI
Se	63.0%	54.8%	71.2%
Sp	86.0%	80.0%	92.0%
PPV	82.3%	74.9%	89.7%
NPV	69.3%	62.1%	76.4%

5) CA-125, all malignancy vs all benign, threshold = 35 u/mL

	Dis+	Dis-	Tot
T+	136	41	177
T-	90	130	220
Tot	226	171	397

	Value	Lower 95% CI	Upper 95% CI
Se	60.0%	53.6%	66.4%
Sp	76.0%	69.6%	82.4%
PPV	76.8%	70.5%	83.0%
NPV	59.0%	52.5%	65.5%

6) CA-125, all malignancy vs all benign, threshold = 57 u/mL

	Dis+	Dis-	Tot
T+	118	27	145
T-	108	144	252
Tot	226	171	397

	Value	Lower 95% CI	Upper 95% CI
Se	52.0%	45.5%	58.5%
Sp	84.0%	78.5%	89.5%
PPV	81.1%	74.7%	87.5%
NPV	57.0%	50.9%	63.1%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

7) CA 15-3, all malignancy vs all benign, threshold = 26 u/mL

	Dis+	Dis-	Tot
T+	106	15	122
T-	120	156	275
Tot	226	171	397

	Value	Lower 95% CI	Upper 95% CI
Se	47.0%	40.5%	53.5%
Sp	91.0%	86.7%	95.3%
PPV	87.3%	81.4%	93.3%
NPV	56.5%	50.7%	62.4%

8) CA-724, all malignancy vs all benign, threshold = 3.5 U/mL

	Dis+	Dis-	Tot
T+	118	27	145
T-	108	144	252
Tot	226	171	397

	Value	Lower 95% CI	Upper 95% CI
Se	52.0%	45.5%	58.5%
Sp	84.0%	78.5%	89.5%
PPV	81.1%	74.7%	87.5%
NPV	57.0%	50.9%	63.1%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Schutter, Kenemans, Sohn, et al., 1994 **#940**	**Geographical location:** Amsterdam, The Netherlands; and Heidelberg, Koln, Wurzburg, and Ulm, Germany **Dates:** Nov 1990-Dec 1992 **Size of population:** 276; 48 excluded (13 45 years old or less, 16 not amenorrheic for 12 months, 4 no path dx, 14 additional malignancy, 10 indeterminate pelvic exam, 3 no preop CA-125), for total of 228 **Other** Multicenter, prospective, patients referred for pelvic mass **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Se/Sp; logistic regression **Blinding:** No **Definition of positive**	**Age:** Mean: 63 Range: 45-88 **Menopausal status (n [%]):** 100% postmenopausal **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** 45 years old Amenorrheic > 12 months Schedule for surgery with tissue diagnosis No history of bilateral oophorectomy or additional cancer **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 199 (87.3%) **Detected by imaging (n [%]):** 28 (12.7%) **Combination (n [%]):** NR **Additional data used for diagnosis:** Stage distribution: I: 16/72 (22.2%) II: 7/72 (9.7%) III: 34/72 (47.2%) IV: 15/72 (20.8%)	Borderline tumors (n = 6) not included 1) CA-125 > 35 U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	68	25	93
T-	27	102	129
Tot	95	127	222 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	71.6%	62.5%	80.6%
Sp	80.3%	73.4%	87.2%
PPV	73.1%	64.1%	82.1%
NPV	79.1%	72.0%	86.1% | 2) Ultrasound: Finkler score > 7 (as described) |	Dis+	Dis-	Tot
---	---	---	---
T+	84	46	130
T-	11	81	92
Tot	95	127	222 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	88.4%	82.0%	94.9%
Sp	63.8%	55.4%	72.1%
PPV	64.6%	56.4%	72.8%
NPV	88.0%	81.4%	94.7%		**Comments:** --Clinical presentation not described (symptoms vs. no symptoms) **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: + (95% CI's given) Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	and negative on screening test: Ultrasound: Positive if multiseptated or irregular cystic mass "consistent in appearance with ovarian tumor; or presence of ascites CA-125: > 35 U/mL				
Schutter, Sohn, Kristen, et al., 1998 **#730**	**Geographical location:** Germany **Dates:** NR **Size of population:** 155 women **Other** Unclear **Reference standard:** Histology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** For pelvic exam – No CA-125 – Yes CA-72-4 - ? (mostly in gastric CA) they set own cutoff point here US - Yes **Statistical tests used:** Chi-square ROC Logistic regression **Blinding:** No	**Age:** Mean: 63 Median: 61 Range: 45-88 **Menopausal status (n [%]):** Post (> 55): 155 (100%) **Race/ethnicity (n [%]):** NR (assume 100% white) **Risk factors (n [%]):** NR **Inclusion criteria:** NR (see citation 25) **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) PE Dis+ Dis- Tot T+ 54 24 78 T- 5 68 73 Tot 59 92 151 Value / Lower 95% CI / Upper 95% CI Se 91.5% 84.4% 98.6% Sp 73.9% 64.9% 82.9% PPV 69.2% 59.0% 79.5% NPV 93.2% 87.4% 98.9% 2) US Dis+ Dis- Tot T+ 51 28 79 T- 8 64 72 Tot 59 92 151 Value / Lower 95% CI / Upper 95% CI Se 86.4% 77.7% 95.2% Sp 69.6% 60.2% 79.0% PPV 64.6% 54.0% 75.1% NPV 88.9% 81.6% 96.1% 3) CA-125 ≥ 35 U/ml Dis+ Dis- Tot T+ 41 15 56 T- 18 77 95 Tot 59 92 151	**Comments:** --Borderline tumors (n = 4) omitted from 2x2 tables by authors --Note "semi-quantitative parameters" of Finkler scoring system --Finkler scoring system? variable **Quality assessment:** Reference standard: + Verification bias: - unclear how many negatives didn't have surgery Test reliability/variability: + for PE, U/S, CA-125 Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test: Table 1 Exam – clinical impression of malignancy US – Finkler Score 7-10 CA-125 - ≥ 35 U/ml				
Scoutt, McCarthy, Lange, et al., 1994 **#4530**	**Geographical location:** New Haven, CT **Dates:** 1988-1990 **Size of population:** 103 patients with 121 masses; data provided on 120 **Other** Case series **Reference standard:** Surgery **Reference standard applied to all test negatives?:** No; 11 with "classic" leiomyomas, 2 with "no mass' **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Se/Sp **Blinding:** Yes **Definition of positive**	**Age:** Median: 40 Range: 2-87 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Data provided for mass, not patient MRI, malignant vs. benign; threshold for positive MR = indeterminate or malignant Dis+ / Dis- / Tot T+: 21 / 11 / 32 T-: 1 / 87 / 88 Tot: 22 / 98 / 120 Value (95% CI: Lower – Upper) Se: 95.5% (86.8% – 100.0%) Sp: 88.8% (82.5% – 95.0%) PPV: 65.6% (49.2% – 82.1%) NPV: 98.9% (96.6% – 100.0%)	**Comments:** --Clinical presentation not described --Length of followup not described **Quality assessment:** Reference standard: - Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	and negative on screening test: Malignancy: solid or highly complex, mural nodules, septations > 3 mm, ascites. Indeterminate if not definitely benign or malignant				
Sengoku, Satoh, Saitoh, et al., 1994 **#4390**	**Geographical location:** Asahikawa, Japan **Dates:** Apr 1991-May 1992 **Size of population:** 28 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Discussed **Statistical tests used:** Se/Sp, t-test **Blinding:** No **Definition of positive and negative on screening test:** Ultrsound morphology	**Age:** Mean (SD): 49 (15) Range: 19-80 **Menopausal status (n [%]):** Pre (< 45): 17 (60.7%) Post (> 55): 11 (39.3%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 35 U/ml, benign vs malignant Dis+ / Dis- / Tot T+: 13 / 3 / 16 T-: 3 / 9 / 12 Tot: 16 / 12 / 28 Value / Lower 95% CI / Upper 95% CI Se: 81.3% / 62.1% / 100.0% Sp: 75.0% / 50.5% / 99.5% PPV: 81.3% / 62.1% / 100.0% NPV: 75.0% / 50.5% / 99.5% 2) Ultrasound morphology Dis+ / Dis- / Tot T+: 13 / 4 / 17 T-: 3 / 8 / 11 Tot: 16 / 12 / 28 Value / Lower 95% CI / Upper 95% CI Se: 81.3% / 62.1% / 100.0% Sp: 66.7% / 40.0% / 93.3% PPV: 76.5% / 56.3% / 96.6% NPV: 72.7% / 46.4% / 99.0% 3) Pulsatility index < 1.5 Dis+ / Dis- / Tot T+: 13 / 1 / 14 T-: 3 / 11 / 14	**Comments:** --Clinical presentation not described --Small numbers --No description of where PI < 1.5 comes from **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding:- Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Sassone system (referenced, not described) Pulsatility index: < 1.5 CA-125> > 35			Tot 16 12 28 Value / Lower 95% CI / Upper 95% CI Se 81.3% 62.1% 100.0% Sp 91.7% 76.0% 100.0% PPV 92.9% 79.4% 100.0% NPV 78.6% 57.1% 100.0%	
Shabana and Onsrud, 1994 **#4400**	**Geographical location:** Trondheim, Norway **Dates:** NR **Size of population:** 85 women **Other – Case control** Three groups compared in terms of serum markers:33 with ovarian Ca, 26 with benign pelvic masses, and 26 with normal pelvis **Reference standard:** Histopathology for all but those with "normal pelvis" which was assessed at laparoscopy **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes for CA-125 **Statistical tests used:** Se, Sp, ROC	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 25 U/ml Dis+ / Dis- / Tot T+ 26 2 28 T- 4 24 28 Tot 30 26 56 Value / Lower 95% CI / Upper 95% CI Se 86.7% 74.5% 98.8% Sp 92.3% 82.1% 100.0% PPV 92.9% 83.3% 100.0% NPV 85.7% 72.8% 98.7%	**Comments:** --Not prospective data --Only 26 with benign pelvic masses included in analysis – no discussion of this in text. --CA-125 cutoff – 25 – not discussed in text --Unclear how the three groups chosen **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Blinding: NR **Definition of positive and negative on screening test:** CA-125 > 25 U/ml				
Siegel, Dehdashti, Mutch, et al., 2003 **#1960**	**Geographical location:** St Louis, MO; Houston, TX; Indianapolis, IN; **Dates:** NR **Size of population:** 35; 2 did not undergo surgery **Other** Validation **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se/Sp **Blinding:** Yes **Definition of positive and negative on screening test:** "Focal increased uptake of radiotracer at sites not	**Age:** Mean: 55 Range: 31-78 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Results provided only for newly diagnosed pelvic masses (n = 26) Borderline excluded (faint uptake by 1 reader, none by other) Dis+ / Dis- / Tot T+: 7 / 4 / 11 T-: 0 / 14 / 14 Tot: 7 / 18 / 25 Value (Lower 95% CI, Upper 95% CI) Se: 100.0% (57.1%, 100.0%) Sp: 77.8% (58.6%, 97.0%) PPV: 63.6% (35.2%, 92.1%) NPV: 100.0% (78.6%, 100.0%) Specificity increased to 82% when re-read with clinical information available	**Comments:** --Clinical presentation not described --Not a commonly used test modality **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	corresponding to obvious normal anatomic structures"				
Smikle, Lunt, and Hankins, 1995 **#6290**	**Geographical location:** USA **Dates:** Jun 1990 – Aug 1992 **Size of population:** 195 **Retrosceptive case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Fisher exact test **Blinding:** NR **Definition of positive and negative on screening test:** CA-125 >= 35 U/ml	**Age:** Those with benign lesions: mean age – 45 (14.7) Those with malignant lesions: mean – 56.9 (13.9) **Menopausal status (n [%]):** Not clear for those for whom CA-125 was available (n = 100) 50 were > 50 years old and 50 were <=50 **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Chart review of all cases with operative reports of "rule-out malignancy" "pelvic mass" "adnexal mass" in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 >= 35 U/ml in postmenopausal women (age >50) Dis+ 5 Tot T+ 11 26 37 T- 2 32 34 Tot 13 58 71 Value / Lower 95% CI / Upper 95% CI Se 84.6% 65.0% 100.0% Sp 55.2% 42.4% 68.0% PPV 29.7% 15.0% 44.5% NPV 94.1% 86.2% 100.0% 2) CA-125 in women <= 50 Dis+ 5 Tot T+ 2 12 14 T- 0 36 36 Tot 2 48 50 Value / Lower 95% CI / Upper 95% CI Se 100.0% -50.0% 100.0% Sp 75.0% 62.8% 87.3% PPV 14.3% 0.0% 32.6% NPV 100.0% 91.7% 100.0% 3) CA-125 for all women combined Dis+ 5 Tot T+ 13 17 30 T- 2 68 70 Tot 15 85 100 Value / Lower 95% CI / Upper 95% CI Se 86.7% 69.5% 100.0% Sp 80.0% 71.5% 88.5%	**Comments:** --Borderline tumors grouped in with malignant --Of 195 charts identified, only 100 had CA-125 levels for analysis --State that platelet count failed to distinguish benign from malignant, however, data not reported such that could be included in meta-analysis --Not recorded whether TVUS or abdominal US done **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: +/- Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				PPV 43.3% 25.6% 61.1% NPV 97.1% 93.2% 100.0%	
Sohaib, Mills, Sahdev, et al., 2005 **#7430**	**Geographical location:** UK **Dates:** NR **Size of population:** 89, however only data for 72 **Prospective case series** NR **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** ROC curves Student's t-test Mann-Whitney U test Chi square **Blinding:** Yes **Definition of positive and negative on screening test:** US – morphology and RI and PI were noted and a "subjective assessment was made as to whether each mass was benign or malignant" and then	**Age:** Mean: 53 Range: 19-86 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Referral to hospital for adnexal mass in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US Dis+ Dis- Tot T+ 29 26 55 T- 0 17 17 Tot 29 43 72 Value Lower 95% CI Upper 95% CI Se 100.0% 89.7% 100.0% Sp 39.5% 24.9% 54.1% PPV 52.7% 39.5% 65.9% NPV 100.0% 82.4% 100.0% 2) MRI Dis+ Dis- Tot T+ 28 7 35 T- 1 36 37 Tot 29 43 72 Value Lower 95% CI Upper 95% CI Se 96.6% 90.0% 100.0% Sp 83.7% 72.7% 94.7% PPV 80.0% 66.7% 93.3% NPV 97.3% 92.1% 100.0%	**Comments:** --Borderline tumors grouped in with malignant --Se, Sp reported for CA-125 , however, no mention of what cutpoint used was in article, and not enough raw data for extraction – authors do state, however, that that MRI performed better than US in cases with normal levels of CA-125. - TVUS done in 65/72 cases – unable to stratify results **Quality assessment:** Reference standard:+ Verification bias:+ Test reliability/variability: + Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	assigned to groups: 1) benign, 2) probably benign, 3) possibly malignant, 4) probably malignant, 5) malignant For Se and Sp calcs, groups 1 and 2 were evaluated together as were 3, 4, 5. MRI – classified into malignant or benign by radiologist "impression" and into same 5 categories.				
Soper, Hunter, Daly, et al., 1990 **#6590**	**Geographical location:** Durham, NC **Dates:** Jan 1985-Jan 1986 **Size of population:** 100 women **Other** Single center Prospective series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes (reference) **Statistical tests used:** NR **Blinding:** Assays done after	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** "Diagnostic laparotomy for pelvic mass" **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125—All malignancies, > 35 U/mL as threshold Dis+ Dis- Tot T+ 38 14 52 T- 16 32 48 Tot 54 46 100 Value Lower 95% CI Upper 95% CI Se 70.4% 58.2% 82.5% Sp 69.6% 56.3% 82.9% PPV 73.1% 61.0% 85.1% NPV 66.7% 53.3% 80.0% 2) CA-125—ovarian cancer, > 35 U/mL as threshold Dis+ Dis- Tot T+ 38 22 60 T- 4 36 40 Tot 42 58 100 Value Lower 95% CI Upper 95% CI Se 90.5% 81.6% 99.4% Sp 62.1% 49.6% 74.6% PPV 63.3% 51.1% 75.5% NPV 90.0% 80.7% 99.3%	Comments: --Unclear if this represents all patients with adnexal mass --Clinical presentation not described --Borderline tumors included with malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: N/AS

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	surgery **Definition of positive and negative on screening test:**			3) CA-125—All malignancies, > 65 U/ml as threshold Dis+ Dis- Tot T+ 42 8 50 T- 12 38 50 Tot 54 46 100 Value / Lower 95% CI / Upper 95% CI Se 77.8% 66.7% 88.9% Sp 82.6% 71.7% 93.6% PPV 84.0% 73.8% 94.2% NPV 76.0% 64.2% 87.8% 4) CA-125---Ovarian cancer, > 65 U/ml as threshold Dis+ Dis- Tot T+ 35 15 50 T- 7 43 50 Tot 42 58 100 Value / Lower 95% CI / Upper 95% CI Se 83.3% 72.1% 94.6% Sp 74.1% 62.9% 85.4% PPV 70.0% 57.3% 82.7% NPV 86.0% 76.4% 95.6% 5) TAG-72---Ovarian cancer Dis+ Dis- Tot T+ 21 4 25 T- 21 54 75 Tot 42 58 100 Value / Lower 95% CI / Upper 95% CI Se 50.0% 34.9% 65.1% Sp 93.1% 86.6% 99.6% PPV 84.0% 69.6% 98.4% NPV 72.0% 61.8% 82.2%	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				6) CA-15-3—Ovarian cancer Dis+ Dis- Tot T+ 23 11 34 T- 19 47 66 Tot 42 58 100 Value / Lower 95% CI / Upper 95% CI Se 54.8% 39.7% 69.8% Sp 81.0% 70.9% 91.1% PPV 67.6% 51.9% 83.4% NPV 71.2% 60.3% 82.1% Results not reported separately, but CA-125 > 65 sensitivity of 79%, specificity of 86% in patients > 50.	
Stein, Laifer-Narin, Johnson, et al., 1995 **#4280**	**Geographical location:** Los Angeles, CA **Dates:** Jul 1992-Feb 1993 **Size of population:** 161 patients, 170 masses **Other** Single center series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** References provided **Statistical tests used:**	**Age:** NR **Menopausal status (n [%]):** Pre (< 45):114 (70.8%) Post (> 55): 39 (24.2%) 8 (4.9%) post hysterectomy **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** Premenopausal with simple or hemorrhagic cysts that resolved on followup, or if examined on any day other than days 3-10 of menstrual cycle;	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Ultrasound morphology (transvaginal in all but 23) Dis+ Dis- Tot T+ 46 47 93 T- 1 76 77 Tot 47 123 170 Value / Lower 95% CI / Upper 95% CI Se 97.9% 93.7% 100.0% Sp 61.8% 53.2% 70.4% PPV 49.5% 39.3% 59.6% NPV 98.7% 96.2% 100.0% 2) Color Doppler-internal flow within solid component or septation Dis+ Dis- Tot T+ 36 38 74 T- 11 85 96 Tot 47 123 170 Value / Lower 95% CI / Upper 95% CI Se 76.6% 64.5% 88.7%	**Comments:** --Denominator should include those excluded because of findings --Results not presented by menopausal status --Analysis, 2x2 tables done for masses not patients **Quality assessment:** Reference standard: Verification bias: - Test reliability/variability: + Sample size: + Statistical tests:+ Blinding: - Definition of +/- on screening test:

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Mann-Whitney **Blinding:** Unclear if radiologists were blinded to clinical data **Definition of positive and negative on screening test:** PI < 1.0 RI < 0.4 Morphology: suggestive of malignancy if complex cystic with solid mural nodules, complex cystic with thick septations ≥ 3 mm, predominately solid, or solid	corpus luteum cyst on pathology; masses determined to arise from uterus on ultrasound		Sp 69.1% 60.9% 77.3% PPV 48.6% 37.3% 60.0% NPV 88.5% 82.2% 94.9% 3) Spectral Doppler—pulsatility index < 1.0 Dis+ Dis- Tot T+ 31 42 73 T- 16 81 97 Tot 47 123 170 Value Lower 95% CI Upper 95% CI Se 67.0% 53.6% 80.4% Sp 66.0% 57.6% 74.4% PPV 42.5% 31.1% 53.8% NPV 83.5% 76.1% 90.9% 4) Spectral Doppler—resistive index < 0.4 Dis+ Dis- Tot T+ 11 12 23 T- 36 111 147 Tot 47 123 170 Value Lower 95% CI Upper 95% CI Se 23.4% 11.3% 35.5% Sp 90.2% 85.0% 95.5% PPV 47.8% 27.4% 68.2% NPV 75.5% 68.6% 82.5%	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Strigini, Gadducci, Del Bravo, et al., 1996 **#4000**	**Geographical location:** Pisa, Italy **Dates:** Jan 1993-June 1994 **Size of population:** 109 women **Other** Case series **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Mann-Whitney U test, chi-square **Blinding:** Unclear **Definition of positive and negative on screening test:** Ultrasound: solid, irregular structure, thick septae, irregular margins Doppler: PI < 1 CA-125: > 65 U/mL	**Age:** Median: 43 Range: 18-80 **Menopausal status (n [%]):** Pre (< 45): 75 (69%) Post (> 55): 34 (31%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Not clearly described Consecutive patients scheduled for laparotomy for adnexal mass in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** "Most" premenopausal women followed over several menstrual cycles to rule out functional cyst	1) Ultrasound: Premenopausal |	Dis+	Dis-	Tot
---	---	---	---
T+	6	2	8
T-	0	67	67
Tot	6	69	75 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	40.0%	100.0%
Sp	97.1%	93.1%	100.0%
PPV	75.0%	45.0%	100.0%
NPV	100.0%	92.0%	100.0% | 2) Ultrasound: Postmenopausal |	Dis+	Dis-	Tot
---	---	---	---
T+	10	2	12
T-	3	19	22
Tot	13	21	34 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	76.9%	54.0%	99.8%
Sp	90.5%	77.9%	100.0%
PPV	83.3%	62.2%	100.0%
NPV	86.4%	72.0%	100.0% | 3) Ultrasound: Combined pre and post |	Dis+	Dis-	Tot
---	---	---	---
T+	16	4	20
T-	3	86	89
Tot	19	90	109 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	84.2%	67.8%	100.0%
Sp	95.6%	91.3%	99.8%
PPV	80.0%	62.5%	97.5%
NPV	96.6%	92.9%	100.0%	 4) Doppler: Premenopausal	**Comments:** --Clinical presentation not described --Unclear how many excluded by resolution on serial scans --US criteria descriptive – not reproducible? – no discussion of variability of interpretation **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: - Statistical tests:+ Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	5	19	24
T-	1	50	51
Tot	6	69	75

	Value	Lower 95% CI	Upper 95% CI
Se	83.3%	53.5%	100%
Sp	72.5%	61.9%	83.0%
PPV	20.8%	4.6%	37.1%
NPV	98.0%	94.2%	100%

5) Doppler: Postmenopausal

	Dis+	Dis-	Tot
T+	11	4	15
T-	2	17	19
Tot	13	21	34

	Value	Lower 95% CI	Upper 95% CI
Se	84.6%	65.0%	100%
Sp	81.0%	64.2%	97.7%
PPV	73.3%	51.0%	95.7%
NPV	89.5%	75.7%	100%

6) Doppler: pre and post combined

	Dis+	Dis-	Tot
T+	16	23	39
T-	3	67	70
Tot	19	90	109

	Value	Lower 95% CI	Upper 95% CI
Se	84.2%	67.8%	100.0%
Sp	74.4%	65.4%	83.5%
PPV	41.0%	25.6%	56.5%
NPV	95.7%	91.0%	100.0%

7) CA-125: Premenopausal

	Dis+	Dis-	Tot
T+	3	7	10
T-	3	62	65

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Tot	6	69	75

	Value	Lower 95% CI	Upper 95% CI
Se	50.0%	10.0%	90.0%
Sp	89.9%	82.7%	97.0%
PPV	30.0%	1.6%	58.4%
NPV	95.4%	90.3%	100.0%

8) CA-125: Postmenopausal

	Dis+	Dis-	Tot
T+	8	2	10
T-	5	19	24
Tot	13	21	34

	Value	Lower 95% CI	Upper 95% CI
Se	61.5%	35.1%	88.0%
Sp	90.5%	77.9%	100%
PPV	80.0%	55.2%	100%
NPV	79.2%	62.9%	95.4%

9) CA-125: combined

	Dis+	Dis-	Tot
T+	77	9	86
T-	8	81	89
Tot	85	90	175

	Value	Lower 95% CI	Upper 95% CI
Se	90.6%	84.4%	96.8%
Sp	90.0%	83.8%	96.2%
PPV	89.5%	83.1%	96.0%
NPV	91.0%	85.1%	97.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Szpurek, Moszyniki, and Sajdak, 2004 **#1530**	**Geographical location:** Poland **Dates:** NR **Size of population:** 464 women **Case series** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Se, Sp, ROC **Blinding:** No **Definition of positive and negative on screening test:** DS index ≥ 4 Doppler Subjective Index	**Age:** Pre-menopause: Mean (SD): 37.4 (9.7) Post-menopause: Mean (SD): 63.4 (8.7) **Menopausal status (n [%]):** Pre: 363 (78%) Post: 101 (22%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women undergoing surgery for ovarian mass who had had Doppler u/s done at the hospital **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** **DS index** 1 point given for each of the following: Number of vessels (≥ 5) Location of vessels (in septae, in papillae or solid parts) Arrangement of vessels (Irregular, random) Shape of velocity waves (smooth; low waveform amplitude) Presence of protodiastolic notch (absent)	1) Doppler index (cutoff DS ≥ 4) postmenopausal women only (n = 101) Dis+ / Dis- / Tot T+: 68 / 0 / 68 T-: 6 / 27 / 33 Tot: 74 / 27 / 101 Value (Lower 95% CI, Upper 95% CI) Se: 91.9% (85.7%, 98.1%) Sp: 100.0% (88.9%, 100.0%) PPV: 100.0% (95.6%, 100.0%) NPV: 81.8% (68.7%, 95.0%) AUC = 0.9775 2) Doppler index (cutoff DS ≥ 4) all women (n = 464) Dis+ / Dis- / Tot T+: 143 / 20 / 163 T-: 22 / 279 / 301 Tot: 165 / 299 / 464 Value (Lower 95% CI, Upper 95% CI) Se: 86.7% (81.5%, 91.9%) Sp: 93.3% (90.5%, 96.1%) PPV: 87.7% (82.7%, 92.8%) NPV: 92.7% (89.8%, 95.6%) AUC = 0.9315 3) Doppler index (cutoff DS ≥ 4) premenopause Dis+ / Dis- / Tot T+: 75 / 20 / 95 T-: 16 / 252 / 268 Tot: 91 / 272 / 363 Value (Lower 95% CI, Upper 95% CI) Se: 82.4% (74.6%, 90.2%) Sp: 92.6% (89.5%, 95.7%)	**Comments:** --Cut off point for DS calculated for this study **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				PPV 78.9% 70.7% 87.1% NPV 94.0% 91.2% 96.9%	
Tailor, Bourne, Campbell, et al., 2003 **#1970**	**Geographical location:** UK **Dates:** NR **Size of population:** 2500 screened **Screening study** **Reference standard:** Histopathology or repeat US at between 12 weeks and six months depending on characteristic of first US and individual family history **Reference standard applied to all test negatives?:** No **Test reliability established?:** No for US Yes for CA-125 **Statistical tests used:** Se Sp, Posterior, prior oddss **Blinding:** Not described – but prospective study **Definition of positive and negative on screening test:** US – considered	**Age:** Mean: 48 Range: 17-78 **Menopausal status (n [%]):** Pre (< 45): 1629 (65%) Post (> 55): 644 (26%) And 227(9%) were post hysterectomy **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** See below **Inclusion criteria:** Self-referred women in time frame for screening Women with at least 1 relative with ovarian cancer and another with another cancer **Exclusion criteria:** NR	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** 0 (0%) **Detected by imaging (n [%]):** 0 **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US for the first screening episode Dis+ 5 Tot T+ 6 76 82 T- 1 2417 2418 Tot 7 2493 2500 Value Lower 95% CI Upper 95% CI Se 85.7% 59.8% 100.0% Sp 97.0% 96.3% 97.6% PPV 7.3% 1.7% 13.0% NPV 100.0% 99.9% 100.0% 2) US for the second screening episode (n = 998) Dis+ 5 Tot T+ 3 11 14 T- 0 984 984 Tot 3 995 998 Value Lower 95% CI Upper 95% CI Se 100.0% 0.0% 100.0% Sp 98.9% 98.2% 99.5% PPV 21.4% 0.0% 42.9% NPV 100.0% 99.7% 100.0% 3) for >= 3rd screen episode (n = 733) Dis+ 5 Tot T+ 2 6 8 T- 0 725 725 Tot 2 731 733 Value Lower 95% CI Upper 95% CI Se 100.0% -50.0% 100.0% Sp 99.2% 98.5% 99.8%	**Comments:** --Borderline grouped in with malignant --Good data on family history specific risk (however, data organized by screen events, not by patient) --Menopause not defined --Unable to calculate 2x2 table for CA-125 as don't know either the true negative rate or the N --Subjective morphologic criteria used --Inclusion criteria shifted through study period --TVUS only **Quality assessment:** Reference standard: - Verification bias: +/- Test reliability/variability: - Sample size: + Statistical tests: + Blinding: + Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	abnormal is: "ovary enlarged (>95th centile), a cyst was detected, and areas of hypo- or hyperechogenicity were seen which were inconsistent with normalphysiology" CA-125 – range used			PPV 25.0% 0.0% 55.0% NPV 100.0% 99.6% 100.0%	
Takac, 1998 **#3240**	**Geographical location:** Maribor, Slovenia **Dates:** Jan 1994-Dec 1995 **Size of population:** 120 women **Other** Case series **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes; referenced and discussed **Statistical tests used:** Mann-Whitney **Blinding:** Not mentioned **Definition of positive and negative on screening test:** ROC done on varying	**Age:** Overall not reported; mean age of patients with benign masses 42 (13.5), malignant 53.3 (17.1) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Adnexal mass that had surgery **Exclusion criteria:** Not undergoing surgery	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 120 (100%) had mass on examination in clinic—but unclear what initial presentation **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Premenopausal US in follicular phase	1) Resistive index ≤ 0.4 Dis+ Dis- Tot T+ 32 2 34 T- 7 76 83 Tot 39 78 117 Value Lower 95% CI Upper 95% CI Se 82.1% 70.0% 94.1% Sp 97.4% 93.9% 100.0% PPV 94.1% 86.2% 100.0% NPV 91.6% 85.6% 97.5% 3 patients with absent vascularization (2 benign, 1 malignant) excluded from analysis in paper; reported Se = 82%, Sp = 97%	**Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: + Sample size: - Statistical tests: - Blinding: + Definition of +/- on screening test:

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	definitions for resistive index Both highest and lowest measured RI used in calculating final RI				
Tanir, Ozalp, Yalcin, et al., 2003 **#1850**	**Geographical location:** Eskisehir, Turkey **Dates:** Aug 1991-Sept 2002 **Size of population:** 63 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** ROC **Blinding:** Yes **Definition of positive and negative on screening test:** Varied cutoff for VEGF, CA-125	**Age:** Reported separately by diagnosis: Non-neoplastic 39.0 (2.0) Benign 42.2 (5.2) Malignant 56.9 (4.2) **Menopausal status (n [%]):** Pre (< 45): 40 (63%) Post (> 55): 23 (37%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Ultrasound (including Doppler)	1) VEGF—threshold 68.7 Dis+ / Dis- / Tot T+ 11 6 17 T- 1 44 45 Tot 12 50 62 Value / Lower 95% CI / Upper 95% CI Se 92.0% 76.7% 100.0% Sp 88.0% 79.0% 97.0% PPV 64.7% 42.0% 87.4% NPV 97.8% 93.5% 100.0% 2) CA-125 ≥ 37 U/mL Dis+ / Dis- / Tot T+ 11 23 34 T- 1 27 28 Tot 12 50 62 Value / Lower 95% CI / Upper 95% CI Se 92.0% 76.7% 100.0% Sp 54.0% 40.2% 67.8% PPV 32.4% 16.6% 48.1% NPV 96.4% 89.6% 100.0%	**Comments:** --Clinical presentation not described -- Borderline tumors grouped in with malignant --Unable to calculate 2x2 tables by menopausal status, but AUC's given: Premenopausal VEGF: AUC 0.938 CA-125: AUC 0.769 Postmenopausal VEGF: AUC 0.902 CA-125: 0.873 Reported Se Sp inconsistent with reported LR+ and LR-, 2x2 tables uncertain. **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: _ Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Tay and Chua, 1994 **#4450**	**Geographical location:** Singapore **Dates:** Oct 1991 – Apr 1992 **Size of population:** 105 women **Other** Prospective study of patients admitted to single center with mass for surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** **Definition of positive and negative on screening test:** CA-125 > 35 U/mL	**Age:** Mean (SD): 38.6 (12.9) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients admitted to hospital with diagnosis of "ovarian cysts" for elective surgery. **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** 105 (100%) **Additional data used for diagnosis:** NR	1) Serum CA-125 > 35 U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	8	21	29
T-	1	74	75
Tot	9	95	104 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	88.9%	68.4%	100.0%
Sp	77.9%	69.6%	86.2%
PPV	27.6%	11.3%	43.9%
NPV	98.7%	96.1%	100.0%		**Comments:** --Unclear how patients chosen ("cysts") --Study compares salivary CA-125 with urine and serum --There appears to be an error in reported statistics – not all of Se, Sp, PPV, and NPV can be correct. "In this study the false positive rate was 22%..." suggests that the Sp reported as 79.2 % was in error and should be 78% (or 77.9% as shown in our 2x2 table abstraction). **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - not discussed Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Tekay and Jouppila, 1992 **#10970**	**Geographical location:** Finland **Dates:** NR **Size of population:** 72 **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp Mann-Whitney U test **Blinding:** Not mentioned but prospective **Definition of positive and negative on screening test:** RI <= 0.6 or 0.5	**Age:** Premenopausal – mean 37 range 17-50 Post menopausal – mean age 60 range 42-74 **Menopausal status (n [%]):** Pre (< 45): 46 (63.9%) Post (> 55): 26(36.1%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Referral to hospital for adnexal mass who underwent surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI <=0.6 Dis+ / Dis- / Tot T+: 9 / 17 / 26 T-: 2 / 44 / 46 Tot: 11 / 61 / 72 Value (95% CI) Se: 82.0% (59.3% – 100.0%) Sp: 72.0% (60.7% – 83.3%) PPV: 34.6% (16.3% – 52.9%) NPV: 95.7% (89.8% – 100.0%) 2) RI <+ 0.5 Dis+ / Dis- / Tot T+: 5 / 7 / 12 T-: 6 / 54 / 60 Tot: 11 / 61 / 72 Value (95% CI) Se: 46.0% (16.5% – 75.5%) Sp: 89.0% (81.1% – 96.9%) PPV: 41.7% (13.8% – 69.6%) NPV: 90.0% (82.4% – 97.6%)	**Comments:** --Borderline tumors grouped with malignant --Although data collected on US morphology and PI, only data for RI able to be extracted into 2x2 table --Overlap in PI and RI noted for all malignant and benign tumors --68 examined with TVUS, 4 with abdominal US – unable to stratify results **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Tepper, Lerner-Geva, Altaras, et al., 1995 **#4090**	**Geographical location:** Kfar Saba, Israel **Dates:** 1990-1993 **Size of population:** 217 women **Other** All patients admitted to hospital with mass for surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes, by reference **Statistical tests used:** Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** RI cutoff of < 0.4 (literature standard) compared with 0.47 (study mean) and 0.53 (mean + 2SD)	**Age:** Mean (SD): 44.3 Range: 8-79 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Admitted to hospital for surgery for diagnosis of adnexal mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI (cutoff < 0.4), only considering patients in whom blood flow velocity waveforms could be detected, borderline tumors excluded |	Dis+	Dis-	Tot
---	---	---	---
T+	17	3	20
T-	8	79	87
Tot	25	82	107 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	68.0%	49.7%	86.3%
Sp	96.3%	92.3%	100.0%
PPV	85.0%	69.4%	100.0%
NPV	90.8%	84.7%	96.9% | 2) RI (cutoff < 0.4), borderline tumors counted as malignant |	Dis+	Dis-	Tot
---	---	---	---
T+	19	3	22
T-	33	162	195
Tot	52	165	217 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	36.5%	23.5%	49.6%
Sp	98.2%	96.1%	100.0%
PPV	86.4%	72.0%	100.0%
NPV	83.1%	77.8%	88.3% | 3) RI (cutoff < 0.4) borderline tumors counted as benign |	Dis+	Dis-	Tot
---	---	---	---
T+	17	5	22
T-	21	174	195
Tot	38	179	217 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	44.7%	28.9%	60.5%
Sp	97.2%	94.8%	99.6%		**Comments:** --To translate data into 2x2 table, abstractor had to round figures, so could be introducing error --Unclear inclusion criteria --Interobserver variability not discussed Tumors of low malignant potential are not included in 2x2 tables Authors calculated Se/Sp only on patients in whom blood flow velocity waveforms could be detected: 25/38 (65.8%) malignant tumors 12/14 (85.7%) borderline tumors 82/165 (49.7%) benign tumors **Quality assessment:** Reference standard: + Verification bias: ± Test reliability/variability: - Sample size: - Statistical tests: +/- Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

PPV	77.3%	59.8%	94.8%
NPV	89.2%	84.9%	93.6%

4) RI (cutoff < 0.47)

	Dis+	Dis-	Tot
T+	33	23	56
T-	5	142	147
Tot	38	165	203

	Value	Lower 95% CI	Upper 95% CI
Se	86.8%	76.1%	97.6%
Sp	86.1%	80.8%	91.3%
PPV	58.9%	46.0%	71.8%
NPV	96.6%	93.7%	99.5%

5) RI (cutoff < 0.53)

	Dis+	Dis-	Tot
T+	38	43	81
T-	0	122	122
Tot	38	165	203

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	92.1%	100.0%
Sp	73.9%	67.2%	80.6%
PPV	46.9%	36.0%	57.8%
NPV	100.0%	97.5%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Tian, Zhang, Jiao, et al., 2000 **#2670**	**Geographical location:** Beijing, China **Dates:** April 1996-Nov 1998 **Size of population:** 71 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Chi-square, t-test **Blinding:** Yes; explicitly stated **Definition of positive and negative on screening test:** Grade 1 or 2 uptake of c-99m, no move with peristalsis or volume change of bladder, or signs of intestinal, peritoneal, or lymphatic involvement	**Age:** Benign Range: 15-72 Malignant Range: 25-70 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Mass, scheduled for laparoscopy **Exclusion criteria:** NR	**Symptomatic (n [%]):** "Most" had "no symptoms other than slight abdominal pain" **Detected by exam (n [%]):** 65 (91.5%; unclear if screening exam) **Detected by imaging (n [%]):** 28 (39.4%); MRI or CT **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Tc-99m |	Dis+	Dis-	Tot
---	---	---	---
T+	22	7	29
T-	1	41	42
Tot	23	48	71 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	92.0%	80.9%	100.0%
Sp	54.0%	39.9%	68.1%
PPV	75.9%	60.3%	91.4%
NPV	97.6%	93.0%	100.0% | 2) CA-125 (>35 U/ml) |	Dis+	Dis-	Tot
---	---	---	---
T+	20	12	32
T-	3	18	21
Tot	23	30	53 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	87.0%	73.2%	100.0%
Sp	60.0%	42.5%	77.5%
PPV	62.5%	45.7%	79.3%
NPV	85.7%	70.7%	100.0% | 3) CT-MRI |	Dis+	Dis-	Tot
---	---	---	---
T+	10	8	18
T-	3	7	10
Tot	13	15	28 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	76.9%	54.0%	99.8%
Sp	46.7%	21.4%	71.9%
PPV	55.6%	32.6%	78.5%
NPV	70.0%	41.6%	98.4%		**Comments:** --Small study --Different numbers of patients had different tests --Clinical presentation not reported --Explicitly states that readers blinded to clinical history --No description of CT-MRI diagnostic criteria (unclear if analysis based on outcome of either CT or MRI or both) **Quality assessment:** Reference standard:+ Verification bias: + Test reliability/variability:- Sample size: - Statistical tests: - Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Timmerman, Verrelst, Bourne, et al., 1999 **#5940**	**Geographical location:** UK **Dates:** Aug 1994 – Aug 1996 **Size of population:** 173 women **Other** Consecutive case series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Logistic regression ROC Se Sp **Blinding:** NR - prospective **Definition of positive and negative on screening test:** CA-125 ≥ 35	**Age:** Mean: Pre 40 Post 64 Range: 22-93 **Menopausal status (n [%]):** Pre (< 45): 83 (48.0%) Post (> 55): 90 (52.0%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Prospective patients to get surgery for mass **Exclusion criteria:** No CA-125 preoperatively available	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 ≥ 35 (borderline included as malignant) Dis+ / Dis- / Tot T+: 39 / 23 / 62 T-: 10 / 101 / 111 Tot: 49 / 124 / 173 Value (95% CI) Se: 79.6% (68.3% – 90.9%) Sp: 81.5% (74.7% – 88.3%) PPV: 62.9% (50.9% – 74.9%) NPV: 91.0% (85.7% – 96.3%) 2) CA-125 ≥ 35 (borderline included as benign) Dis+ / Dis- / Tot T+: 34 / 27 / 61 T-: 10 / 102 / 112 Tot: 44 / 129 / 173 Value (95% CI) Se: 77.3% (64.9% – 89.7%) Sp: 79.1% (72.0% – 86.1%) PPV: 55.7% (43.3% – 68.2%) NPV: 91.1% (85.8% – 96.4%)	**Comments:** --Clinical presentation not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: +/- Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Timor-Tritsch, Lerner, Montea-gudo, et al., 1993 **#10840**	**Geographical location:** USA **Dates:** Apr 1991 – May 1992 **Size of population:** 93 patients 115 masses **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** Not described but prospective **Definition of positive and negative on screening test:** Sassone scoring system Used (cutoff not mentioned but assume >= 9) RI < 0.46 PI < 0.62	**Age:** Mean: 43.2 Range: 13-74 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Presenting to hospital with diagnosis of mass in time frame **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Sassone morphologic criteria Dis+ / Dis- / Tot T+: 11 / 9 / 20 T-: 1 / 61 / 62 Tot: 12 / 70 / 82 Value (Lower 95% CI – Upper 95% CI) Se: 93.7% (80.0% – 100.0%) Sp: 87.1% (79.2% – 95.0%) PPV: 55.0% (33.2% – 76.8%) NPV: 98.4% (95.3% – 100.0%) 2) PI < 0.62 Dis+ / Dis- / Tot T+: 11 / 2 / 13 T-: 2 / 68 / 70 Tot: 12 / 70 / 83 Value (Lower 95% CI – Upper 95% CI) Se: 87.5% (68.8% – 100.0%) Sp: 97.4% (93.7% – 100.0%) PPV: 84.6% (65.0% – 100.0%) NPV: 97.1% (93.2% – 100.0%) 3) RI < 0.46 Dis+ / Dis- / Tot T+: 11 / 1 / 12 T-: 1 / 69 / 70 Tot: 12 / 70 / 82 Value (Lower 95% CI – Upper 95% CI) Se: 93.8% (80.2% – 100.0%) Sp: 98.7% (96.0% – 100.0%) PPV: 91.7% (76.0% – 100.0%) NPV: 98.6% (95.8% – 100.0%)	**Comments:** --Excluded from analysis those masses in which no flow could be measured (although none were CA) --Borderline tumors grouped in with malignant --RI and PI cutpoints calculated from results of patient series (predefined cutpoints not used) --Don't mention cutpoint for Sassone's criteria (assume 9 from original article) --Clinical pathway not illuminated - -Although much of the article data was reported for masses not individuals, these 2x2 tables are for individuals --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: +/- Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Tingulstad, Hagen, Skjeldestad, et al., 1996 **#3890**	**Geographical location:** Norway **Dates:** Feb 1992 – Feb 1994 **Size of population:** 173 women **Other** Consecutive patient series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Chi square, Se Sp, ROC Mann-Whitney U test **Blinding:** NR - prospective **Definition of positive and negative on screening test:** CA-125 – range US ≥ 2, where 1 point assigned for each of the following: multilocular cystic lesion solid areas bilateral lesion ascites intra-abdominal ascites	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 82 (47.4%) Post (> 55): 91 (52.6%) (defined as > 12 months amenorrhea or age > 50 if hysterectomy) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age >30 with pelvic mass scheduled for surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 25 U/ml Dis+ / Dis- / Tot T+: 45 / 22 / 67 T-: 11 / 95 / 106 Tot: 56 / 117 / 173 Value (95% CI) Se: 80.0% (69.5% – 90.5%) Sp: 81.0% (73.9% – 88.1%) PPV: 67.2% (55.9% – 78.4%) NPV: 89.6% (83.8% – 95.4%) 2) CA-125 > 50 U/ml Dis+ / Dis- / Tot T+: 41 / 7 / 48 T-: 15 / 110 / 125 Tot: 56 / 117 / 173 Value (95% CI) Se: 73.0% (61.4% – 84.6%) Sp: 94.0% (89.7% – 98.3%) PPV: 85.4% (75.4% – 95.4%) NPV: 88.0% (82.3% – 93.7%) 3) US score > 2 Dis+ / Dis- / Tot T+: 38 / 21 / 59 T-: 18 / 96 / 114 Tot: 56 / 117 / 173 Value (95% CI) Se: 68.0% (55.8% – 80.2%) Sp: 82.0% (75.0% – 89.0%) PPV: 64.4% (52.2% – 76.6%) NPV: 84.2% (77.5% – 90.9%) 4) Menopausal status (postmenopausal = malignant)	**Comments:** --RMI study – no cutoffs in analysis --No description of inter **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - /intraobserver variability with US Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	45	46	91
T-	11	71	82
Tot	56	117	173

	Value	Lower 95% CI	Upper 95% CI
Se	80.4%	70.0%	90.8%
Sp	60.7%	51.8%	69.5%
PPV	49.5%	39.2%	59.7%
NPV	86.6%	79.2%	94.0%

Torres, Derchain, Faundes, et al., 2002

#2170

Geographical location:
Sao Paulo, Brazil

Dates: Jan 1996 – Mar 1998

Size of population:
158 women

Other
Series

Reference standard:
Histopathology

Reference standard applied to all test negatives?:
Yes

Test reliability established?:
Yes

Statistical tests used:
Se, Sp
Regression

Blinding:
NR – but prospective

Age:
NR

Menopausal status (n [%]):
NR

Race/ethnicity (n [%]):
NR

Risk factors (n [%]):
NR

Inclusion criteria:
Pelvic mass who had surgery in time frame

Exclusion criteria:
Lung masses,
Signs of hepatic or intraperitoneal mets

Symptomatic (n [%]):
NR

Detected by exam (n [%]):
NR

Detected by imaging (n [%]):
NR

Combination (n [%]):
NR

Additional data used for diagnosis:
NR

1) CA-125 > 35 U/ml

	Dis+	Dis-	Tot
T+	52	23	75
T-	15	68	83
Tot	67	91	158

	Value	Lower 95% CI	Upper 95% CI
Se	78.0%	68.1%	87.9%
Sp	75.0%	66.1%	83.9%
PPV	69.3%	58.9%	79.8%
NPV	81.9%	73.6%	90.2%

2) US score ≥ 2

	Dis+	Dis-	Tot
T+	55	30	85
T-	12	61	73
Tot	67	91	158

	Value	Lower 95% CI	Upper 95% CI
Se	82.0%	72.8%	91.2%
Sp	67.0%	57.3%	76.7%
PPV	64.7%	54.5%	74.9%
NPV	83.6%	75.1%	92.1%

3) US score ≥ 3

Comments:
--RMI study so no exact cutoffs given, rather 2x2 tables done at various levels
--Unclear where patients referred from or how decision to do surgery done
-No discussion of variability in US

Quality assessment:
Reference standard: +
Verification bias: +
Test reliability/variability: CA-125 + US -
Sample size: -
Statistical tests: +
Blinding: +/- (not discussed but prospective)
Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Definition of positive and negative on screening test:
US score – range from other RMI literature (0-8)
CA-125 - range

	Dis+	Dis-	Tot
T+	50	26	76
T-	17	65	82
Tot	67	91	158

	Value	Lower 95% CI	Upper 95% CI
Se	75.0%	64.6%	85.4%
Sp	71.0%	61.7%	80.3%
PPV	65.8%	55.1%	76.5%
NPV	79.3%	70.5%	88.0%

4) Menopausal status (menopausal = malignant)

	Dis+	Dis-	Tot
T+	49	28	77
T-	18	63	81
Tot	67	91	158

	Value	Lower 95% CI	Upper 95% CI
Se	73.0%	62.4%	83.6%
Sp	69.0%	59.5%	78.5%
PPV	63.6%	52.9%	74.4%
NPV	77.8%	68.7%	86.8%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Troiano, Quedens-Case, and Taylor, 1997 **#3680**	**Geographical location:** New Haven, CT **Dates:** 1991-1996 **Size of population:** 144 women **Other** Case series **Reference standard:** Surgery (45); serial ultrasound (followup 1-6 years) **Reference standard applied to all test negatives?:** See above **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Descriptive **Blinding:** No **Definition of positive and negative on screening test:** Solid components, thickened septation, nodules; Dopplers; no explicit scoring or cutpoint used. CA-125 ≥ 36 U/m;	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 102 (70.8%) Post (> 55): 42 (29.2%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 ≥ 36 U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	17	88	105
T-	3	36	39
Tot	20	124	144 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	85.0%	69.4%	100.0%
Sp	29.0%	21.0%	37.0%
PPV	16.2%	9.1%	23.2%
NPV	92.3%	83.9%	100.0% | 2) CA-125 ≥ 36 U/ml, postmenopausal |	Dis+	Dis-	Tot
---	---	---	---
T+	10	25	35
T-	0	7	7
Tot	10	32	42 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	70.0%	100.0%
Sp	21.9%	7.6%	36.2%
PPV	28.6%	13.6%	43.5%
NPV	100.0%	57.1%	100.0%		**Comments:** --Reported sensitivity/specificity not consistent with data provided in Table 1 of paper --Diagnostic criteria not explicit --Unable to calculate 2x2 table for ultrasound results --Borderline tumors grouped with malignant **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: - -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Valentin, 1997 **#6170**	**Geographical location:** Malmo, Sweden **Dates:** NR **Size of population:** 151 women **Other** Prospective study at university hospital of consecutive patients with adnexal mass scheduled for surgery **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes, but 49 were excluded for technical reasons **Test reliability established?:** **Statistical tests used:** Se, Sp, regression **Blinding:** NR **Definition of positive and negative on screening test:**	**Age:** Mean: 47.8 Range: 20-90 **Menopausal status (n [%]):** Pre (< 45): 92 (62.9%) Post (> 55): 53 (37.1%) 6 women who had undergoing hysterectomy were classified a pre or post menopausal on the basis of age. **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** 49 excluded for: uninterpretable US examinations (12), surgery canceled (replaced by laparoscopy) (5), biopsy with cytologic diagnosis (3), or clinical followup (22), myoma at time of surgery (7)	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) PI (cutoff < 1) |	Dis+	Dis-	Tot
---	---	---	---
T+	19	74	93
T-	4	38	42
Tot	23	112	135 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	82.6%	67.1%	98.1%
Sp	33.9%	25.2%	42.7%
PPV	20.4%	12.2%	28.6%
NPV	90.5%	81.6%	99.4% | 2) Color lakes visible on Doppler |	Dis+	Dis-	Tot
---	---	---	---
T+	21	42	63
T-	3	85	88
Tot	24	127	151 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	87.5%	74.3%	100.0%
Sp	66.9%	58.7%	75.1%
PPV	33.3%	21.7%	45.0%
NPV	96.6%	92.8%	100.0% | 3) Menopausal status (post =T+) |	Dis+	Dis-	Tot
---	---	---	---
T+	19	5	24
T-	37	90	127
Tot	56	95	151 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	33.9%	21.5%	46.3%
Sp	94.7%	90.2%	99.2%
PPV	79.2%	62.9%	95.4%
NPV	70.9%	63.0%	78.8%	 Multiple logistic regression model constructed for multilocular solid tumors –	**Comments:** --Negative surgeries but with diagnosis of myoma were excluded (7) --Unclear why the 5 who had only diagnostic laparoscopy were excluded --LMP tumors lumped in with malignant --Visibility of color lakes is subjective evaluation of Doppler, uncertain reliability **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				results not reported here.	
Valentin, 1999b **#3100**	**Geographical location:** Malmo, Sweden **Dates:** NR **Size of population:** 173 women **Other** Series **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** Chi-square **Blinding:** No **Definition of positive and negative on screening test:** Morphology: solid components Lerner score: ≥ 3 Color score: ≥ 40 or 65 Pulsatility < 1.0 Resistive index < 0.4 Time-averaged max velocity: ≥ 7.2 cm/s Peak systolic velocity: ≥ 14.4 cm/s	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 98 (59.5%) Peri (45-55): Post (> 55): 70 (40.5%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) "Subjective" Morphology alone Dis+ / Dis- / Tot T+: 21 / 6 / 27 T-: 3 / 143 / 146 Tot: 24 / 149 / 173 Value / Lower 95% CI / Upper 95% CI Se: 87.5% / 74.3% / 100.0% Sp: 96.0% / 92.8% / 99.1% PPV: 77.8% / 62.1% / 93.5% NPV: 97.9% / 95.6% / 100.0% 2) Morphology plus Doppler Dis+ / Dis- / Tot T+: 20 / 5 / 25 T-: 4 / 144 / 148 Tot: 24 / 149 / 173 Value / Lower 95% CI / Upper 95% CI Se: 83.3% / 68.4% / 98.2% Sp: 96.6% / 93.8% / 99.5% PPV: 80.0% / 64.3% / 95.7% NPV: 97.3% / 94.7% / 99.9% 3) Lerner score Dis+ / Dis- / Tot T+: 22 / 96 / 118 T-: 2 / 53 / 55 Tot: 24 / 149 / 173 Value / Lower 95% CI / Upper 95% CI Se: 91.7% / 80.6% / 100.0% Sp: 35.6% / 27.9% / 43.3% PPV: 18.6% / 11.6% / 25.7% NPV: 96.4% / 91.4% / 100.0%	**Quality assessment:** Reference standard: - Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + **Quality assessment:** Reference standard: - Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

4) Time-averaged max velocity

	Dis+	Dis-	Tot
T+	24	61	85
T-	0	59	59
Tot	24	120	144

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	87.5%	100.0%
Sp	49.2%	40.2%	58.1%
PPV	28.2%	18.7%	37.8%
NPV	100.0%	94.9%	100.0%

5) Morphology plus peak systolic velocity

	Dis+	Dis-	Tot
T+	19	44	63
T-	5	99	104
Tot	24	143	167

	Value	Lower 95% CI	Upper 95% CI
Se	79.2%	62.9%	95.4%
Sp	69.2%	61.7%	76.8%
PPV	30.2%	18.8%	41.5%
NPV	95.2%	91.1%	99.3%

6) Morphology plus color score

	Dis+	Dis-	Tot
T+	17	39	56
T-	7	106	113
Tot	24	145	169

	Value	Lower 95% CI	Upper 95% CI
Se	70.8%	52.6%	89.0%
Sp	73.1%	65.9%	80.3%
PPV	30.4%	18.3%	42.4%
NPV	93.8%	89.4%	98.2%

7) Morphology

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	23	94	117
T-	1	55	56
Tot	24	149	173

	Value	Lower 95% CI	Upper 95% CI
Se	95.8%	87.8%	100.0%
Sp	36.9%	29.2%	44.7%
PPV	19.7%	12.5%	26.9%
NPV	98.2%	94.7%	100.0%

8) Resistance index

	Dis+	Dis-	Tot
T+	6	7	13
T-	18	113	131
Tot	24	120	144

	Value	Lower 95% CI	Upper 95% CI
Se	25.0%	7.7%	42.3%
Sp	94.2%	90.0%	98.4%
PPV	46.2%	19.1%	73.3%
NPV	86.3%	80.4%	92.2%

9) Pulsatility index

	Dis+	Dis-	Tot
T+	21	81	102
T-	3	39	42
Tot	24	120	144

	Value	Lower 95% CI	Upper 95% CI
Se	87.5%	74.3%	100.0%
Sp	32.5%	24.1%	40.9%
PPV	20.6%	12.7%	28.4%
NPV	92.9%	85.1%	100.0%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Valentin, 2000 **#2760**	**Geographical location:** Malmo, Sweden **Dates:** NR **Size of population:** 173 women **Other** Consecutive patients scheduled for surgery in university hospital **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Lerner's ? PI and RI yes Valentin's Doppler variable - no **Statistical tests used:** Se, Sp **Blinding:** NR – prospective though **Definition of positive and negative on screening test:** Lerner's score [4] >3 RI < 0.4 PI < 1.0	**Age:** Median: pre 37.5 for post 66 Range: 18-88 **Menopausal status (n [%]):** Pre: 98 (56.6%) Post: 70 (40.5%) 4 hysterectomy 1 unknown **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women admitted to hospital for surgery due to pelvic mass during time frame **Exclusion criteria:** 26 excluded because 10 surgery canceled or replaced with cytology, 13 had normal US preop of whom 9 had normal laparoscopy and 4 had normal US followup	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Lerner score ≥ 3 |	Dis+	Dis-	Tot
---	---	---	---
T+	22	53	75
T-	2	96	98
Tot	24	149	173 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	91.7%	80.6%	100.0%
Sp	64.4%	56.7%	72.1%
PPV	29.3%	19.0%	39.6%
NPV	98.0%	95.2%	100.0% | 2) PI < 1.0 |	Dis+	Dis-	Tot
---	---	---	---
T+	21	81	102
T-	3	68	71
Tot	24	149	173 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	87.5%	74.3%	100.0%
Sp	45.6%	37.6%	53.6%
PPV	20.6%	12.7%	28.4%
NPV	95.8%	91.1%	100.0% | 3) RI < 0.4 |	Dis+	Dis-	Tot
---	---	---	---
T+	6	7	13
T-	18	142	160
Tot	24	149	173 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	25.0%	7.7%	42.3%
Sp	95.3%	91.9%	98.7%
PPV	46.2%	19.1%	73.3%
NPV	88.8%	83.9%	93.6%	 Additional analyses for combination of Lerner score ≥3 and Doppler variable	**Comments:** --Undetectable velocity measures classified as "benign" --LMP tumors grouped into malignant --Lerner score not described **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability:- Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
				positive, combined as "both" or "either" = T+	
Valentin, Hagen, Tingulstad, et al., 2001 **#2340**	**Geographical location:** Trondheim, Norway, and Malmo, Sweden **Dates:** NR **Size of population:** 136 **Other** Prospective validation **Reference standard:** Surgery **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Referenced **Statistical tests used:** ROC, logistic regression **Blinding:** No **Definition of positive and negative on screening test:** "Pattern recognition" referenced, but not described in this paper	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Scheduled for surgery for adnexal mass **Exclusion criteria:** Surgery cancelled No histology Death Multiple lesions in same ovary	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) "Pattern recognition" (threshold not defined, but referenced) Dis+ / Dis- / Tot T+: 30 / 9 / 39 T-: 6 / 91 / 97 Tot: 36 / 100 / 136 Value (Lower 95% CI, Upper 95% CI) Se: 83.3% (71.2%, 95.5%) Sp: 91.0% (85.4%, 96.6%) PPV: 76.9% (63.7%, 90.1%) NPV: 93.8% (89.0%, 98.6%)	**Comments:** --For CA-125 – unable to get data for 2x2 table as it is reported in Means only --For US – scoring system broken into parts in tables – unable to reassemble to get 2x2 table for US alone --Threshold given in references 1 and 2 (Valentin et al, Ultrasound Obstet Gynecol 1999;14:273-83, and 1999:14:338-47) --Clinical presentation not described **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: + Sample size: - (wide CI's) Statistical tests: + Blinding: - Definition of +/- on screening test: + (not given here, but referenced)

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
van Nagell Jr., DePriest, Reedy, et al., 2000 **#2730**	**Geographical location:** USA **Dates:** 1987 - 1999 **Size of population:** 14469 180with persistently abnormal screens leading to surgery **Screening study** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** No Normal screens were repeated in one year Abnormal screens were repeated in 4-6 weeks with morphology, CA-125 and Doppler **Test reliability established?:** yes **Statistical tests used:** Chi square Fisher's exact test Kaplan Meier survival curve plotted **Blinding:** Prospective **Definition of positive and negative on screening test:** US – ovarian volume > 20cm3 for	**Age:** Mean (SD): 54.7 (10.7) Range: 25-92 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age >=50 or age >=25 with documented family history of ovarian CA **Exclusion criteria:** NR	**Symptomatic (n [%]):** 0 (0%) **Detected by exam (n [%]):** 0 (0%) **Detected by imaging (n [%]):** 180 (1.2%) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) For the screened population as a whole Dis+ / 5 / Tot T+: 17 / 163 / 180 T-: 4 / 14285 / 14289 Tot: 21 / 14448 / 14469 Value / Lower 95% CI / Upper 95% CI Se: 81.0% / 64.2% / 97.7% Sp: 98.9% / 98.7% / 99.0% PPV: 9.4% / 5.2% / 13.7% NPV: 100.0% / 99.9% / 100.0% 2) Adding the 4 (Table 6) who got CA > 12 months out Dis+ / 5 / Tot T+: 17 / 163 / 180 T-: 8 / 14281 / 14289 Tot: 25 / 14444 / 14469 Value / Lower 95% CI / Upper 95% CI Se: 68.0% / 49.7% / 86.3% Sp: 98.9% / 98.7% / 99.0% PPV: 9.4% / 5.2% / 13.7% NPV: 99.9% / 99.9% / 100.0%	**Comments:** --4 patients developed ovarian CA 12+ months after screen – these were included in the true negative analysis (because they had failed in 12 month followup) --Borderline tumors grouped in with benign --TVUS only **Quality assessment:** Reference standard: - Verification bias: + Test reliability/variability: - Sample size: + Statistical tests: + Blinding: + Definition of +/- on screening test: +/-

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	premenopausal women, and > 10cm3 for postmenopausal women Or any cystic ovarian tumor with a solid or papillary projection into its lumen. For subsequent screening DePriest scoring system used with CA-125 and doppler				
Vasilev, Schlaerth, Campeau, et al., 1988 **#6770**	**Geographical location:** Los Angeles, CA **Dates:** Mar 1984-Feb 1986 **Size of population:** 182 women **Other** Nonconsecutive case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** NR **Blinding:** Yes **Definition of positive**	**Age:** NR **Menopausal status (n [%]):** ≤ 50: 152 (83.1%) >50: 31 (17.0%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Preoperative diagnosis of pelvic mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) All patients CA-125 ≥ 35 U/ml Dis+ Dis- Tot T+ 14 37 51 T- 4 128 132 Tot 18 165 183 Value / Lower 95% CI / Upper 95% CI Se 77.8% 58.6% 97.0% Sp 77.6% 71.2% 83.9% PPV 27.5% 15.2% 39.7% NPV 97.0% 94.0% 99.9% 2) Age ≤ 50 Dis+ Dis- Tot T+ 6 35 41 T- 0 111 111 Tot 6 146 152 Value / Lower 95% CI / Upper 95% CI Se 100.0% 50.0% 100.0% Sp 76.0% 69.1% 83.0% PPV 14.6% 3.8% 25.5% NPV 100.0% 97.3% 100.0% 3) Age > 50	**Comments:** --Unclear what implications of "nonconsecutive" are—how many patients not included --Text description does not match data provided in Table 3 --Borderline grouped with malignant **Quality assessment:** Reference standard: - Verification bias: - Test reliability/variability:+ Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	and negative on screening test: > 35 u/ML				

	Dis+	Dis-	Tot
T+	8	2	10
T-	4	17	21
Tot	12	19	31

	Value	Lower 95% CI	Upper 95% CI
Se	66.7%	40.0%	93.3%
Sp	89.5%	75.7%	100.0%
PPV	80.0%	55.2%	100.0%
NPV	81.0%	64.2%	97.7%

Vuento, Pirhonen, Makinen, et al., 1995

#4070

Geographical location: Turku, Finland

Dates: NR

Size of population: 1846 in initial pool; 1364 (74%) consented

Screening study

Reference standard: Cases reported to Finnish Cancer Registry over 2 ½ year followup; true negative considered no cancer within 1 year

Reference standard applied to all test negatives?: Yes

Test reliability established?: Not referenced or described

Statistical tests used: Chi-square

Age: Mean (SD): 59 Range: 56-61

Menopausal status (n [%]): Post (> 55): 1364 (100%)

Race/ethnicity (n [%]): NR

Risk factors (n [%]): Family history: 376 (27.6%)

Inclusion criteria: Presenting for mammography

Exclusion criteria: NR

Symptomatic (n [%]): 100% asymptomatic

Detected by exam (n [%]): NR

Detected by imaging (n [%]): NR

Combination (n [%]): NR

Additional data used for diagnosis: NR

1) Combined US morphology and Doppler

	Dis+	Dis-	Tot
T+	1	159	160
T-	0	1204	1204
Tot	1	1363	1364

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	0.0%	100.0%
Sp	88.3%	86.6%	90.0%
PPV	0.6%	0.0%	1.8%
NPV	100.0%	99.8%	100.0%

Positive case was borderline
2 cancers, 1 Stage 1A and 1 Stage III, within 2 years

2) Combined US morphology and Doppler, results of followup US in 160 women with abnormal screening ultrasound on first screen

	Dis+	Dis-	Tot
T+	1	27	28
T-	0	132	132
Tot	1	159	160

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	0.0%	100.0%

Comments:
--Population based study
--Some potential for verification bias, but registry capture reasonable alternative

Quality assessment:
Reference standard: +
Verification bias: +
Test reliability/variability:-
Sample size: +
Statistical tests: +
Blinding: +
Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	Blinding: Yes **Definition of positive and negative on screening test:** Negative: not visualized, volume less than 8 cc, uniformly hypoechogenic; PI < 1.0 – lowest measured value used US – Fleischer criteria			Sp 83.0% 77.2% 88.9% PPV 3.6% 0.0% 10.4% NPV 100.0% 97.7% 100.0%	
Wakahara, Kikkawa, Nawa, et al., 2001 **#2370**	**Geographical location:** Nagoya, Japan **Dates:** 1994-1999 **Size of population:** 292 women **Other** Case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced or discussed **Statistical tests used:** Se/Sp **Blinding:** Yes **Definition of positive and negative on**	**Age:** Mean (SD): 40.3 Range: 11-79 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Presence of adnexal mass, scheduled for surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Ultrasound—Including borderline as cancer Dis+ Dis- Tot T+ **69** **37** 106 T- **15** **171** 186 Tot 84 208 292 Value Lower 95% CI Upper 95% CI Se 82.1% 74.0% 90.3% Sp 82.2% 77.0% 87.4% PPV 65.1% 56.0% 74.2% NPV 91.9% 88.0% 95.8% 2) Ultrasound—classifying borderline as benign Dis+ Dis- Tot T+ **57** **49** 106 T- **9** **177** 186 Tot 66 226 292 Value Lower 95% CI Upper 95% CI Se 86.4% 78.1% 94.6% Sp 78.3% 72.9% 83.7% PPV 53.8% 44.3% 63.3% NPV 95.2% 92.1% 98.2% 3) CA-125 > 35 U/ml; borderline classified	**Comments:** --Clinical presentation not described --Not stratified by menopausal status **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	screening test: Positive ultrasound: cyst with irregular solid component or thickened septum, or solid mass. Ultrasound interpretation by gyn oncologist, not radiologist CA-125: 35 U/mL CA-19-9: 37 IU/mL CA-72-4: 4 U/mL			as malignant Dis+ Dis- Tot T+ **62** **22** 84 T- **77** **131** 208 Tot 139 153 292 Value Lower 95% CI Upper 95% CI Se 44.6% 36.3% 52.9% Sp 85.6% 80.1% 91.2% PPV 73.8% 64.4% 83.2% NPV 63.0% 56.4% 69.5% 4) CA-125 > 35 U/ml; borderline classified as benign Dis+ Dis- Tot T+ **51** **88** 139 T- **15** **138** 153 Tot 66 226 292 Value Lower 95% CI Upper 95% CI Se 77.3% 67.2% 87.4% Sp 61.1% 54.7% 67.4% PPV 36.7% 28.7% 44.7% NPV 90.2% 85.5% 94.9% 5) CA 19-9 > 37 IU/ml, borderline classified as malignant Dis+ Dis- Tot T+ **29** **78** 107 T- **55** **127** 182 Tot 84 205 289 Value Lower 95% CI Upper 95% CI Se 34.5% 24.4% 44.7% Sp 62.0% 55.3% 68.6% PPV 27.1% 18.7% 35.5% NPV 69.8% 63.1% 76.5% 6) CA 19-9 >37 IU/ml, Borderline classified	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

as benign

	Dis+	Dis-	Tot
T+	24	83	107
T-	42	140	182
Tot	66	223	289

	Value	Lower 95% CI	Upper 95% CI
Se	36.4%	24.8%	48.0%
Sp	62.8%	56.4%	69.1%
PPV	22.4%	14.5%	30.3%
NPV	76.9%	70.8%	83.0%

7) CA-72-4 > 4 U/ml, borderline classified as malignant

	Dis+	Dis-	Tot
T+	23	18	41
T-	28	104	132
Tot	51	122	173

	Value	Lower 95% CI	Upper 95% CI
Se	45.1%	31.4%	58.8%
Sp	85.2%	79.0%	91.5%
PPV	56.1%	40.9%	71.3%
NPV	78.8%	71.8%	85.8%

8) CA-72-4 > 4 U/ml, borderline classified as benign

	Dis+	Dis-	Tot
T+	20	21	41
T-	20	116	136
Tot	40	137	177

	Value	Lower 95% CI	Upper 95% CI
Se	50.0%	34.5%	65.5%
Sp	84.7%	78.6%	90.7%
PPV	48.8%	33.5%	64.1%
NPV	85.3%	79.3%	91.2%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Weiner, Thaler, Beck, et al., 1992 **#6480**	**Geographical location:** Haifa, Israel **Dates:** NR **Size of population:** 62 women **Other** Prospective surgical candidates in single center **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** No – 9 had "functional cysts" which had regressed on followup US **Test reliability established?:** CA-125 – yes PI +/- **Statistical tests used:** Se, Sp **Blinding:** NR – prospective study **Definition of positive and negative on screening test:** CA-125 > 35 U/ml PI < 1.0 US – {7} – not described	**Age:** Range: 20-69 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women referred to US department for suspicion of pelvic tumor 62 referred – 9 with cysts that disappeared after 6 weeks **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 > 35 U/ml |	Dis+	Dis-	Tot
---	---	---	---
T+	14	14	28
T-	3	22	25
Tot	17	36	53 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	82.4%	64.2%	100.0%
Sp	61.1%	45.2%	77.0%
PPV	50.0%	31.5%	68.5%
NPV	88.0%	75.3%	100.0% | 2) PI < 1.0 |	Dis+	Dis-	Tot
---	---	---	---
T+	16	1	17
T-	1	35	36
Tot	17	36	53 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	94.1%	82.9%	100.0%
Sp	97.2%	91.9%	100.0%
PPV	94.1%	82.9%	100.0%
NPV	97.2%	91.9%	100.0% | 3) US |	Dis+	Dis-	Tot
---	---	---	---
T+	16	11	27
T-	1	25	26
Tot	17	36	53 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	94.1%	82.9%	100.0%
Sp	69.4%	54.4%	84.5%
PPV	59.3%	40.7%	77.8%
NPV	96.2%	88.8%	100.0%		**Comments:** --LMP grouped with malignant for analysis --Classification of sonographic findings unclear [7] reference doesn't illuminate scoring system --PPV for CA-125 reported in article as 45% (this may be typographical error) --Borderline cystadenocarcinomas are classified as malignant in 2x2 tables **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + for CA-125 and PI - for US Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + for CA-125 and PI - for US

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Woolas, Conaway, Xu, et al., 1995 **#4140**	**Geographical location:** Durham, NC, and London, UK **Dates:** NR **Size of population:** 429 **Other** Series from two hospitals **Reference standard:** Surgery, pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** ROC, logistic modeling **Blinding:** Not described **Definition of positive and negative on screening test:** CA-125: 35 u/ML M-CSF: 3.1 ng/ml OVX 1: 12.1 U/mL LASA: 200.0 mg/ml CA 15-3: 32.0 u/ml CA-72-4: 3.8 u/ml CA 19-9: 39.0 u/ml CA 54/61: 20.0 u/ml	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** All patients with "clinically detected masses" **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** 148 primary ovarian cancers 26% Stage I 3% Stage II 55% Stage III 16% Stage IV	1) CA-125 > 35 U/ml Dis+ / Dis- / Tot T+: 150 / 55 / 205 T-: 42 / 182 / 224 Tot: 192 / 237 / 429 Value (95% CI) Se: 78.1% (72.3% - 83.9%) Sp: 76.8% (71.4% - 82.2%) PPV: 73.2% (67.1% - 79.2%) NPV: 81.3% (76.1% - 86.4%) 2) M-CSF Dis+ / Dis- / Tot T+: 127 / 57 / 184 T-: 65 / 180 / 245 Tot: 192 / 237 / 429 Value (95% CI) Se: 66.2% (59.5% - 72.9%) Sp: 76.0% (70.6% - 81.4%) PPV: 69.0% (62.3% - 75.7%) NPV: 73.5% (67.9% - 79.0%) 3) OVX1 Dis+ / Dis- / Tot T+: 77 / 41 / 118 T-: 115 / 196 / 311 Tot: 192 / 237 / 429 Value (95% CI) Se: 40.1% (33.2% - 47.0%) Sp: 82.7% (77.9% - 87.5%) PPV: 65.3% (56.7% - 73.8%) NPV: 63.0% (57.7% - 68.4%) 4) LASA	**Comments:** --Clinical presentation not described- --Spectrum of disease (cancer stage distribution) adequately described --Unclear how subjects selected (random, consecutive, or to approximated distribution of disease --Not stratified by age or menopausal status --Prevalence of cancer higher than would be expected in general population **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: + Statistical tests: + Blinding: + Definition of +/- on screening test: + **Comments:** --Clinical presentation not described- --Spectrum of disease (cancer stage distribution) adequately described --Unclear how subjects selected (random, consecutive, or to approximated distribution of disease --Not stratified by age or menopausal status --Prevalence of cancer higher than would be expected in general population **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: + Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	100	29	129
T-	92	208	300
Tot	192	237	429

	Value	Lower 95% CI	Upper 95% CI
Se	52.1%	45.0%	59.2%
Sp	87.8%	83.6%	92.0%
PPV	77.5%	70.3%	84.7%
NPV	69.3%	64.1%	74.6%

5) CA 15-3

	Dis+	Dis-	Tot
T+	120	32	152
T-	72	205	277
Tot	192	237	429

	Value	Lower 95% CI	Upper 95% CI
Se	62.5%	55.7%	69.3%
Sp	86.5%	82.1%	90.9%
PPV	78.9%	72.5%	85.4%
NPV	74.0%	68.8%	79.2%

6) CA-72-4

	Dis+	Dis-	Tot
T+	104	20	124
T-	88	217	305
Tot	192	237	429

	Value	Lower 95% CI	Upper 95% CI
Se	54.2%	47.2%	61.2%
Sp	91.6%	88.1%	95.1%
PPV	83.9%	77.4%	90.3%
NPV	71.1%	66.1%	76.2%

7) CA 19-9

	Dis+	Dis-	Tot
T+	46	28	74
T-	146	209	355

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

Tot	192	237	429

	Value	Lower 95% CI	Upper 95% CI
Se	24.0%	18.0%	30.0%
Sp	88.2%	84.1%	92.3%
PPV	62.2%	51.1%	73.2%
NPV	58.9%	53.8%	64.0%

8) CA 54/61

	Dis+	Dis-	Tot
T+	100	28	128
T-	92	209	301
Tot	192	237	429

	Value	Lower 95% CI	Upper 95% CI
Se	52.1%	45.0%	59.2%
Sp	88.2%	84.1%	92.3%
PPV	78.1%	71.0%	85.3%
NPV	69.4%	64.2%	74.6%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Wu, Lee, Chen, et al., 1994 **#10900**	**Geographical location:** Taiwan **Dates:** Jul 1990 – Dec 1993 **Size of population:** 410 **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Chi square Fisher exact Student t test Linear regression **Blinding:** Not mentioned (prospective) **Definition of positive and negative on screening test:** Positive: RI > 0.4	**Age:** Range: 11-81 **Menopausal status (n [%]):** NR: **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Suspected malignancy on the basis of suspicious ultrasound **Exclusion criteria:** Patients examined during luteal phase	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RI < 0.4 |	Dis+	Dis-	Tot
---	---	---	---
T+	70	8	78
T-	33	299	332
Tot	103	307	410 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	68.0%	59.0%	77.0%
Sp	97.4%	95.6%	99.2%
PPV	89.7%	83.0%	96.5%
NPV	90.1%	86.8%	93.3%		**Comments:** --Clinical pathway not described --Combined TVUS and abdominal US (no N for each reported) unable to stratify **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: Sample size: Statistical tests: Blinding: Definition of +/- on screening test:

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Yamashita, Hatanaka, Torashima, et al., 1997 **#6120**	**Geographical location:** Japan **Dates:** NR **Size of population:** 104 women **Other** MRI series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** LR **Blinding:** Yes for MRI, independent interpretation by 3 radiologists - surgical and pathological reports were not available **Definition of positive and negative on screening test:** MRI evaluated for 1) size; 2) bilaterality; 3) wall structure; 4) internal architectures; 5) presence of thick (< 3 mm) septa; 6) signal intensity; 7) ascites	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Ovarian masses and MRI to further investigate indeterminate ultrasound **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 100% ultrasound **Combination (n [%]):** NR **Additional data used for diagnosis:** **LR model** Logit(y)= -2.7 + 0.4 (tumor size) + 1.8 (ascites) + 1.4 (bilateral) + 0.5 (complex internal architecture) + 1.6 (solid or irregular wall structure) Where tumor size (cm), other variables scored 1if present; 0 if absent. Best discrimination obtained in test set with cutoff our 0.49.	1) LR model (validation set) cutoff = 0.49 Dis+ / Dis- / Tot T+: 21 / 4 / 25 T-: 7 / 43 / 50 Tot: 28 / 47 / 75 Value (Lower 95% CI, Upper 95% CI) Se: 75.0% (59.0%, 91.0%) Sp: 91.5% (83.5%, 99.5%) PPV: 84.0% (69.6%, 98.4%) NPV: 86.0% (76.4%, 95.6%)	**Comments:** --Low malignant potential treated as malignant in this analysis. --LR model fit 5 variables to 50 positive cases (at limit of 1:10 rule of thumb) --2x2 table reported to validation doesn't match articles reported Se, Sp. Our abstraction based on 2x2 table. --Validation in separate data set. **Quality assessment:** Reference standard: + Verification bias: Test reliability/variability: - Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: -, cutoff was data driven.

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Yamashita, Torashima, Hatanaka, et al., 1995 **#4290**	**Geographical location:** Kumamoto, Japan **Dates:** NR **Size of population:** 72 women 80 masses **Other** Consecutive case series **Reference standard:** Surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** **Inter-rater reliability measured** **Statistical tests used:** ROC, kappa **Blinding:** Yes **Definition of positive and negative on screening test:** Benign if 3 of 4 criteria: • Diameter ≤ 4 cm • Entirely cystic • Lesion wall < 3mm • No internal structure Malignancy if 2 of 5 criteria: • Diameter > 4 cm • Wall or septum >	**Age:** Mean (SD): 43 Range: 13-74 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR; presumably 100% Asian **Risk factors (n [%]):** NR **Inclusion criteria:** Pelvic mass, scheduled for surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Transvaginal ultrasound, benign vs malignant, all readings by 5 radiologists (results presented by lesion, not patient; 19 patients had malignancy, 61 benign lesions) Dis+ / Dis- / Tot T+: 17 / 10 / 27 T-: 2 / 51 / 53 Tot: 19 / 61 / 80 Value / Lower 95% CI / Upper 95% CI Se: 89.0% / 74.9% / 100.0% Sp: 84.0% / 74.8% / 93.2% PPV: 63.0% / 44.7% / 81.2% NPV: 96.2% / 91.1% / 100.0% 2) Preconstrast MRI, benign vs malignant Dis+ / Dis- / Tot T+: 15 / 4 / 19 T-: 4 / 57 / 61 Tot: 19 / 61 / 80 Value / Lower 95% CI / Upper 95% CI Se: 78.0% / 59.4% / 96.6% Sp: 93.0% / 86.6% / 99.4% PPV: 78.9% / 60.6% / 97.3% NPV: 93.4% / 87.2% / 99.7% 3) Contrast-enhanced MRI, benign vs malignant Dis+ / Dis- / Tot T+: 17 / 4 / 21 T-: 2 / 57 / 59 Tot: 19 / 61 / 80 Value / Lower 95% CI / Upper 95% CI Se: 91.0% / 78.1% / 100.0% Sp: 93.0% / 86.6% / 99.4% PPV: 81.0% / 64.2% / 97.7%	**Comments:** --Observer variability measured Kappa 0.71 for pre-contrast MRI, 0.73 for contrast enhanced MRI, 0.62 for transvaginal ultrasound --Se/Sp calculated by lesion, not patient—CI's subsequently smaller --Not stratified by age, menopausal status --Clinical presentation not described LMP tumors grouped in with malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: + Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results				Comments/Quality Scoring
	3mm • Nodularity, vegetations, or large solid component • Necrosis or hemorrhage in lesion And/or if any 1 of the following criteria: • Involvement of adjacent organs or pelvic sidewall • Peritoneal, mesenteric, or omental lesions • Ascites • Adenopathy Borderline tumors classified as malignant			NPV	96.6%	92.0%	100.0%	

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Zanetta, Vergani, and Lissoni, 1994 **#10850**	**Geographical location:** Milan, Italy **Dates:** May 1992 – May 1993 **Size of population:** 80 **Case series** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp **Blinding:** Not mentioned (prospective) **Definition of positive and negative on screening test:** US morphology – Sassone's criteria Doppler: PI>=1.0 RI range investigated (0.50, 0.56, 0.60) CA-125 >30U/ml	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Surgical referral to hospital secondary to mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** 71(88.8%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 5(6.3%) (infertility work up imaging) **Combination (n [%]):** NR **Additional data used for diagnosis:** 4 (5%) referred for abdominal enlargement but were asymptomatic	1) US Sassone Dis+ / Dis- / Tot T+: 32 / 11 / 43 T-: 1 / 36 / 37 Tot: 33 / 47 / 80 Value (Lower 95% CI – Upper 95% CI) Se: 97.0% (91.2% – 100.0%) Sp: 76.0% (63.8% – 88.2%) PPV: 74.4% (61.4% – 87.5%) NPV: 97.3% (92.1% – 100.0%) 2) CA 126 > 30 Dis+ / Dis- / Tot T+: 30 / 19 / 49 T-: 3 / 28 / 31 Tot: 33 / 47 / 80 Value (Lower 95% CI – Upper 95% CI) Se: 91.0% (81.2% – 100.0%) Sp: 60.0% (46.0% – 74.0%) PPV: 61.2% (47.6% – 74.9%) NPV: 90.3% (79.9% – 100.0%) 3) PI 1.0 Dis+ / Dis- / Tot T+: 32 / 6 / 38 T-: 1 / 41 / 42 Tot: 33 / 47 / 80 Value (Lower 95% CI – Upper 95% CI) Se: 97.0% (91.2% – 100.0%) Sp: 87.0% (77.4% – 96.6%) PPV: 84.2% (72.6% – 95.8%) NPV: 97.6% (93.0% – 100.0%) 4) RI 0.60	**Comments:** --N = 80 however, n = 78 for all the cases that they report --Borderline grouped with malignant --Overlap in both RI and PI noted in malignant and benign lesions --TVUS only **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: +/- Definition of +/- on screening test: +

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	30	7	37
T-	3	40	43
Tot	33	47	80

	Value	Lower 95% CI	Upper 95% CI
Se	91.0%	81.2%	100.0%
Sp	85.0%	74.8%	95.2%
PPV	81.1%	68.5%	93.7%
NPV	93.0%	85.4%	100.0%

5) RI 0.56

	Dis+	Dis-	Tot
T+	28	4	32
T-	5	43	48
Tot	33	47	80

	Value	Lower 95% CI	Upper 95% CI
Se	85.0%	72.8%	97.2%
Sp	91.0%	82.8%	99.2%
PPV	87.5%	76.0%	99.0%
NPV	89.6%	80.9%	98.2%

6) RI 0.50

	Dis+	Dis-	Tot
T+	22	2	24
T-	11	45	56
Tot	33	47	80

	Value	Lower 95% CI	Upper 95% CI
Se	67.0%	51.0%	83.0%
Sp	96.0%	90.4%	100.0%
PPV	91.7%	80.6%	100.0%
NPV	80.4%	70.0%	90.8%

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Zhang, Barnhill, Zhang, et al., 1999 **#3020**	**Geographical location:** London, England **Dates:** NR **Size of population:** 429 women **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** **Statistical tests used:** Se, Sp, ROC **Blinding:** NR **Definition of positive and negative on screening test:** Artificial Neural Network (ANN), cutoff 0.5 Ca 125 > 35U/ml	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Pelvic mass/had pathology **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	2) CA-125 > 35 U/ml Dis+ / Dis- / Tot T+: 75 / 29 / 104 T-: 6 / 57 / 63 Tot: 81 / 86 / 167 Value (95% CI) Se: 92.6% (86.9% – 98.3%) Sp: 66.3% (56.3% – 76.3%) PPV: 72.1% (63.5% – 80.7%) NPV: 90.5% (83.2% – 97.7%)	**Comments:** --Training set originally contained non-epithelial ovarian cancer and other cancers as well as non-ovarian benign conditions but the ANN couldn't be trained so they narrowed the training set to epithelial ovarian malignancy vs known benign ovarian conditions. This may be less applicable to real life than leaving all the other conditions in as well. **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: -

Evidence Table 3 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Zimmer, Tepper, and Akselrod, 2003 **#1740**	**Geographical location:** **Dates:** NR **Size of population:** 28 images; number of patients not described **Other** Development of quantitative analytic method for ultrasound images; initial validation **Reference standard:** Presumably surgery/pathology **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not described **Statistical tests used:** Not described **Blinding:** Yes **Definition of positive and negative on screening test:** Algorithm based on lesion size, structure, turbidity, amount of solid material	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** NR **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Benign vs malignant, from 28 preoperative images Dis+ / Dis- / Tot T+: 20 / 0 / 20 T-: 5 / 3 / 8 Tot: 25 / 3 / 28 Value (Lower 95% CI, Upper 95% CI) Se: 80.0% (64.3%, 95.7%) Sp: 100.0% (0.0%, 100.0%) PPV: 100.0% (85.0%, 100.0%) NPV: 37.5% (4.0%, 71.0%)	**Comments:** --Clinical presentation not described --Spectrum of disease not described **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: - Blinding: + (computerized) Definition of +/- on screening test: +

Evidence Table 4: *Question 4: What is the accuracy of explicit scoring systems which incorporate various combinations of imaging findings, patient risk factors, and/or CA-125 levels for detecting malignancy? Have these scoring systems been applied to a population of women before laparoscopy or laparotomy*?

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Adonakis, Para-skevaidis, Tsiga, et al., 1996 **#810**	**Geographical location:** Ioannina, Greece **Dates:** Mar 1991-Jun 1993 **Size of population:** 2000 screened **Screening study** **Reference standard:** Histology or followup (at least 1 visit with CA-125 1 year later) **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, PPV **Blinding:** No **Definition of positive and negative on screening test:** See Scoring column	**Age:** Mean: 58.1 Range: 45-80 **Menopausal status (n [%]):** Pre: 1302 Peri: 293 Post: 405 **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age ≥ 45 years "without any evidence of adnexal pathology" **Exclusion criteria:** Prior history of ovarian cancer, any other malignancy, bilateral salpingo-oophorectomy or ascites	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 59 115 with "ambiguous" exam **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) TVUS criteria for abnormal include any of the following: a) abnormal morphology – hyper- or hypo-echoic; b) irregular "outline;" c) volume > 18 ml for premenopausal and > 9 ml for postmenopausal women 2) CA-125 > 35 U/ml 3) Pelvic examination abnormal when there is a palpable adnexal mass	1) PE + CA-125 ("ambiguous" PE considered abnormal; no histological verification of test negatives) Dis+ / Dis- / Tot T+: 3 / 177 / 180 T-: 0 / 1820 / 1820 Tot: 3 / 1997 / 2000 Value (Lower 95% CI – Upper 95% CI) Se: 100.0% (0.0% – 100.0%) Sp: 91.1% (89.9% – 92.4%) PPV: 1.7% (0.0% – 3.5%) NPV: 100.0% (99.8% – 100.0%) 2) PE + CA-125 ("ambiguous" PE considered normal; no histological verification of 145 TVUS test negatives) Dis+ / Dis- / Tot T+: 2 / 63 / 65 T-: 1 / 114 / 115 Tot: 3 / 177 / 180 Value (Lower 95% CI – Upper 95% CI) Se: 66.7% (13.3% – 100.0%) Sp: 64.4% (57.4% – 71.5%) PPV: 3.1% (0.0% – 7.3%) NPV: 99.1% (97.4% – 100.0%) 3) PE + CA-125 +TVUS ("ambiguous" PE considered normal; no histological verification of 145 TVUS test negatives) *(see next page)*	**Comments:** --2x2 tables constructed from Table 2 and data reported in text; not able to reproduce the Se, Sp and PPV reported in Table 3 of manuscript --Only one followup visit was required for patients with negative screening – some patients who subsequently developed cancer could have been missed **Quality assessment:** Reference standard: + Verification bias: - (only one followup visit was required for test negatives who did not have surgery) Test reliability/variability: - Sample size: + Statistical tests: - (not enough data given to reproduce results reported) Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	3	68	71
T-	0	109	109
Tot	3	177	180

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	0.0%	100.0%
Sp	61.6%	54.4%	68.7%
PPV	4.2%	0.0%	8.9%
NPV	100.0%	97.2%	100.0%

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Alcazar, Errasti, Zornoza, et al., 1999 **#3110**	**Geographical location:** Pamplona, Spain **Dates:** Jan 1995-Feb 1998 **Size of population:** 94 women of 480 women screened **Registry** Retrospective single-institution series **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, ROC curves/AUC **Blinding:** No **Definition of positive and negative on screening test:** Yes	**Age:** Mean: 47.4 Range: 17-79 **Menopausal status (n [%]):** Pre: 52 (55.3%) Post: 42 (44.7%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Diagnosed as having an adnexal mass and sonographically suspicious findings; transvaginal color Doppler and CA-125 level before surgery; and definitive histopathological diagnosis **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** All 100% **Combination (n [%]):** NR **Additional data used for diagnosis:** Morphological evaluation including presence of at least one of: Multilocularity, gross septations (> 3 mm), gross papillary projections (> 3 mm), solid wall nodules, irregular borders, solid mass or ascites	1) CA-125 ≥ 35 U/ml 2) RI ≤ 0.4	1) Combination of RI ≤ 0.4 and CA-125 ≥ 35 U/ml Dis+ / Dis- / Tot T+: 42 / 1 / 43 T-: 14 / 37 / 51 Tot: 56 / 38 / 94 Value (Lower 95% CI, Upper 95% CI) Se: 75.0% (63.7%, 86.3%) Sp: 97.4% (92.3%, 100.0%) PPV: 97.7% (93.2%, 100.0%) NPV: 72.5% (60.3%, 84.8%)	**Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Andersen, Knudsen, Rix, et al., 2003 **#1810**	**Geographical location:** Northern Jutland, Denmark **Dates:** Jul 1999-Aug 2001 **Size of population:** 447 **Registry** **Reference standard:** Histology or repeat exam every 2-3 months **Reference standard applied to all test negatives?:** No; in 45 cases, patients did not have histology and were observed at 2- to 3-month intervals. No patients were lost to followup. **Statistical tests used:** Se, Sp **Blinding:** No **Definition of positive and negative on screening test:** Well defined, including cutoff value of 200 RMI2 = U x M x CA-125 (Tingulstad, Hagen, Skjeldestad, et al., 1996 [#3890])	**Age:** Range: 30 or older **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** 30 years of age or older; pelvic mass identified in region **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** All had palpable or US demonstrated pelvic mass **Additional data used for diagnosis:** Vaginal ultrasound Menopausal status CA-125	1) CA-125 Absolute value of serum CA-125 (IMX, Abbott Labs) 2) Ultrasound score (U; transvaginal) based on presence of the following factors: solid areas (1) multilocularity (1) bilaterality (1) ascites (1) extraovarian tumors (1) If total ≥ 2, then U = 4 If total < 2, then U = 1 3) Menopausal score (M): Postmenopausal (> 1 year of amenorrhea) = 4 Premenopausal = 1 RMI2 = U x M x CA-125 (Tingulstad, Hagen, Skjeldestad, et al., 1996 [#3890])	1) RMI2 (among all patients; 45 of whom had clinical followup in lieu of histopathological diagnosis) Test + = RMI2 ≥ 200 Test - = RMI2 < 200 |	Dis+	Dis-	Tot
---	---	---	---
T+	72	37	109
T-	30	308	338
Tot	102	345	447 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	70.6%	61.7%	79.4%
Sp	89.3%	86.0%	92.5%
PPV	66.1%	57.2%	74.9%
NPV	91.1%	88.1%	94.2% | 2) RMI2 data limited only to surgically treated patients (all verified with histopathology) |	Dis+	Dis-	Tot
---	---	---	---
T+	72	37	109
T-	30	263	293
Tot	102	300	402 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	70.6%	61.7%	79.4%
Sp	87.7%	83.9%	91.4%
PPV	66.1%	57.2%	74.9%
NPV	89.8%	86.3%	93.2%	 **Results were reported, but have not been abstracted, for the following:** Analysis that considered Stage 1 ovarian cancer as a “benign” disease resulting in Se 95.5% (63/66) and Sp 87.9% (335/381)	**Comments:** --No data on pre- versus postmenopausal women --Predictive value of ultrasound and CA-125 in postmenopausal women cannot be estimated from this report --Included borderline as malignant category **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: ? (unclear from title of reference: Tingulstad, Hagen, Skjeldestad, et al., 1996 [#3890]) Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: + (this is a validation study)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Andolf, Jorgensen, and Astedt, 1990 **#1200**	**Geographical location:** Lund, Sweden **Dates:** Oct 1984-Jul 1987 **Size of population:** 801 women **Screening study Registry** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** No, women without abnormality or cysts < 20 mm diameter considered normal and not verified. Ascertainment of ovarian cancer in test negative relies on Sweden's public health cancer registry system, which is well validated. **Statistical tests used:** Se, Sp **Blinding:** None **Definition of positive and negative on screening test:** NR	**Age:** Range: 40-70 **Menopausal status (n [%]):** Pre (< 45): 330 (41%) Peri (45-55): 212 (26%) Post (> 55): 259 (32%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women between 40-70 years of age in one of 4 groups: 1) abdominal pain, frequent micturition or irregular bleeding; 2) nulliparity; 3) family history of breast, endometrial or ovarian cancer; 4) previous history of cancer **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 106 total 51 (7.9%) of women with normal US 55 women with abnormal US **Detected by imaging (n [%]):** 163 women total 108 women with normal manual exam 55 women with abnormal manual exam **Combination (n [%]):** 55 women with abnormal exam and US **Additional data used for diagnosis:** None	1) Bimanual PE 2) Abdominal US	1) Either US or PE abnormal Dis+ / Dis- / Tot T+: 8 / 206 / 214 T-: 0 / 587 / 587 Tot: 8 / 793 / 801 Value / Lower 95% CI / Upper 95% CI Se: 100.0% / 62.5% / 100.0% Sp: 74.0% / 71.0% / 77.1% PPV: 3.7% / 1.2% / 6.3% NPV: 100.0% / 99.5% / 100.0% 2) Both US and PE abnormal Dis+ / Dis- / Tot T+: 5 / 50 / 55 T-: 3 / 743 / 746 Tot: 8 / 793 / 801 Value / Lower 95% CI / Upper 95% CI Se: 62.5% / 29.0% / 96.0% Sp: 93.7% / 92.0% / 95.4% PPV: 9.1% / 1.5% / 16.7% NPV: 99.6% / 99.1% / 100.0%	**Comments:** --Results are not stratified by risk group or by menopausal status --The portions of diagnostic evaluation that would normally lead to referral (e.g., symptomatic presentation or positive findings on bimanual pelvic exam) are incorporated into the diagnostic assessment in this study. **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: - Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Asif, Sattar, Dawood, et al., 2004 **#1580**	**Geographical location:** Rawalpindi, Pakistan **Dates:** Jan 2001 to Jan 2002 **Size of population:** 100 women **Surgical case series** **Reference standard:** Histopathological diagnosis **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, PPV, NPV **Blinding:** None **Definition of positive and negative on screening test:** Implicitly defined by use of Jacobs instrument (Jacobs, Oram, Fairbanks, et al., 1990 [#6820]), but analysis considered multiple cutoff values	**Age:** Mean (SD): 41.4 **Menopausal status (n [%]):** Pre (< 45): 56 (56%) Post (> 55): 44 (44%) ≥ 1 year amenorrhea **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Consecutive women admitted to one of 2 military hospitals for elective surgical exploration and resection of proven ovarian mass **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 Absolute value of serum CA-125 by solid phase two-site chemiluminescent enzyme immunometric assay using Immulite CA-125 kit (DPC, USA) 2) Ultrasound score (U; transvaginal) based on presence of the following factors: solid areas (1) multilocularity (1) bilaterality (1) ascites (1) extraovarian tumors (1) If total is 0, then U = 0 If total is 1, then U = 1 If total ≥ 2, then U = 3 3) Menopausal score (M): Postmenopausal (> 1 year of amenorrhea) = 3 Premenopausal = 1 RMI1 = U x M x CA-125 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820])	1) RMI1 with cutoff value of 200 Dis+ / Dis- / Tot T+: 47 / 3 / 50 T-: 8 / 42 / 50 Tot: 55 / 45 / 100 Value (Lower 95% CI, Upper 95% CI) Se: 85.0% (75.6%, 94.4%) Sp: 93.0% (85.5%, 100.0%) PPV: 94.0% (87.4%, 100.0%) NPV: 84.0% (73.8%, 94.2%) 2) RMI1 with cutoff value of 125 Dis+ / Dis- / Tot T+: 48 / 5 / 53 T-: 7 / 40 / 47 Tot: 55 / 45 / 100 Value (Lower 95% CI, Upper 95% CI) Se: 87.0% (78.1%, 95.9%) Sp: 88.0% (78.5%, 97.5%) PPV: 90.6% (82.7%, 98.4%) NPV: 85.1% (74.9%, 95.3%) **Results were reported, but have not been abstracted, for the following combinations:** RMI1 with cutoff levels of 25, 50, 75, 100, 150, 175, 190, 250, 300	**Comments:** --No stratification by menopausal status --Incomplete data reporting makes exact numbers in 2x2 tables uncertain: There is some discrepancy with PPV and NPV reported in paper; couldn't find 2x2 cell values to match Se, Sp, PPV and NPV assuming 55 Disease+ and 45 Disease- patients. **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: _ Definition of +/- on screening test: -, did not choose a priori cutoff values Explicit validation method?: + (this was a validation of RMI1)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Aslam, Banerjee, Carr, et al., 2000 **#2690**	**Geographical location:** South London, UK **Dates:** NR **Size of population:** 100 women **Registry** **Reference standard:** Histopathological diagnosis from laparotomy **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, ROC curves/AUC **Blinding:** No **Definition of positive and negative on screening test:** NA	**Age:** Mean: 45.6 Range: 20-78 **Menopausal status (n [%]):** Pre: 63 (63%) Post: 37 (37%) ≥ 1 year amenorrhea or age > 50 if status post hysterectomy **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women with known adnexal masses and due to undergo surgery at one of 3 UK hospitals **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Tailor model LR1 uses age, TAMXV (Doppler), papillary projection score (US): P = 1/(1 + e - z), where z = 0.1273 x age + 0.2794 x TAMXV + 4.4136 x PPS - 14.2046 Cutoff = 50% 2) Timmerman model LR3 uses CA-125, morphologic and demographic data. Z = 0.5948 x menopausal status + 0.0205 x CA-125 + 0.5446 x ascites - 0.762 x unilocularity 1.1606 x smooth + 1.5409 x PPS + 0.7633 x bilateral - 1.0889 P > 50% assumed to be diagnostic of malignancy 3) LR1 + LR2 4) LR1 + LR3 5) LR1 + LR2 + LR3 Where LR2 = Alcazar (1998) model using morphologic features with Doppler blood flow variables Ascites = 1 or 0	1) Tailor model LR1 Dis+ / Dis- / Tot T+: 15 / 5 / 20 T-: 18 / 62 / 80 Tot: 33 / 67 / 100 Value (Lower 95% CI – Upper 95% CI) Se: 45.0% (28.0% – 62.0%) Sp: 93.0% (86.9% – 99.1%) PPV: 75.0% (56.0% – 94.0%) NPV: 77.5% (68.3% – 86.7%) AUC 0.86 (0.77 to 0.94) 2) Timmerman model LR3 Dis+ / Dis- / Tot T+: 24 / 6 / 30 T-: 9 / 61 / 70 Tot: 33 / 67 / 100 Value (Lower 95% CI – Upper 95% CI) Se: 73.0% (57.9% – 88.1%) Sp: 91.0% (84.1% – 97.9%) PPV: 80.0% (65.7% – 94.3%) NPV: 87.1% (79.3% – 95.0%) AUC 0.93 (0.89 to 0.98) 3) LR1+LR2 Dis+ / Dis- / Tot T+: 20 / 5 / 25 T-: 13 / 62 / 75 Tot: 33 / 67 / 100 Value (Lower 95% CI – Upper 95% CI) Se: 61.0% (44.4% – 77.6%) Sp: 93.0% (86.9% – 99.1%) PPV: 80.0% (64.3% – 95.7%) NPV: 82.7% (74.1% – 91.2%) AUC 0.85	**Comments:** --Although study applied reference standard to all test negative women, study included only women already referred for surgery; thus, there is a referral bias in the population. --Borderlines counted as malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: + (study validates 2 previously reported models)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

4) LR1 + LR3

	Dis+	Dis-	Tot
T+	26	4	30
T-	7	63	70
Tot	33	67	100

	Value	Lower 95% CI	Upper 95% CI
Se	79.0%	65.1%	92.9%
Sp	94.0%	88.3%	99.7%
PPV	86.7%	74.5%	98.8%
NPV	90.0%	83.0%	97.0%

AUC 0.95

5) LR2 + LR3

	Dis+	Dis-	Tot
T+	24	11	35
T-	9	56	65
Tot	33	67	100

	Value	Lower 95% CI	Upper 95% CI
Se	73.0%	57.9%	88.1%
Sp	84.0%	75.2%	92.8%
PPV	68.6%	53.2%	84.0%
NPV	86.2%	77.8%	94.6%

AUC 0.88

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Aslam, Tailor, Lawton, et al., 2000 **#2580**	**Geographical location:** South London, UK **Dates:** Jul 1997-Sep 1998 **Size of population:** 61 women **Diagnostic test study** **Reference standard:** Histopathological diagnosis **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp **Blinding:** None **Definition of positive and negative on screening test:** As defined by models used	**Age:** Mean: 46.8 Range: 20-77 **Menopausal status (n [%]):** Pre (< 45): 36 Post (> 55): 25 ≥ 1 year amenorrhea or age > 50 if status post hysterectomy **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women with known adnexal masses due to be admitted for surgery **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RMI1 – Jacobs, Oram, Fairbanks, et al., 1990 (#6820) RMI1 = ultrasound score (U; 0, 1, 3) x menopausal status (M; 1, 3) x serum CA-125 RMI1 > 200 indicates malignancy 1 point each for: Bilateral lesions Multilocular Ascites Solid areas Intraabdominal metastases U = 0 if total is 0 1 if total is 1 and 3 if total is ≥ 2 M: Premenopausal = 1 Postmenopausal = 3 CA-125 in kU/L 2) RMI2 – Tingulstad, Hagen, Skjeldestad, et al., 1996 (#3890) As RMI1 except U = 1 if total is 0 or 1 4 if total ≥ 2; M = 1 or 4 in this model 3) Tailor model [Ref # 17] LR1 uses age, TAMXV (Doppler), PPS (US): P = 1/(1 + e - z), where z = 0.1273 x age + 0.2794 x TAMXV + 4.4136 x PPS - 14.2046	1) RMI1 with cutoff value of 200 |	Dis+	Dis-	Tot
---	---	---	---
T+	17	3	20
T-	6	35	41
Tot	23	38	61 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	73.9%	56.0%	91.9%
Sp	92.1%	83.5%	100.0%
PPV	85.0%	69.4%	100.0%
NPV	85.4%	74.5%	96.2% | 2) RMI2 with cutoff value of 200 |	Dis+	Dis-	Tot
---	---	---	---
T+	17	4	21
T-	6	34	40
Tot	23	38	61 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	73.9%	56.0%	91.9%
Sp	89.5%	79.7%	99.2%
PPV	81.0%	64.2%	97.7%
NPV	85.0%	73.9%	96.1% | 3) Tailor's model with cutoff value of 50% |	Dis+	Dis-	Tot
---	---	---	---
T+	10	3	13
T-	13	35	48
Tot	23	38	61 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	43.5%	23.2%	63.7%
Sp	92.1%	83.5%	100.0%
PPV	76.9%	54.0%	99.8%
NPV	72.9%	60.3%	85.5%		**Comments:** --Although study applied reference standard to all test negative women, study included only women already referred for surgery; thus, there is a referral bias in the population. --Borderlines counted as malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: + (this is a validation of Tailor, Jacobs and Tingulstad)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Balbi, Musone, Menditto, et al., 2001 **#2320**	**Geographical location:** Naples, Italy **Dates:** Jan 1996-Mar 2000 **Size of population:** 92 women **Case series** **Reference standard:** Histopathological diagnosis **Reference standard applied to all test negatives?:** No, 18 women with "clearly benign" masses not verified; 2 patients with "clearly malignant" disease (metastases) also excluded **Statistical tests used:** Se, Sp **Blinding:** None **Definition of positive and negative on screening test:** Defined	**Age:** Range: 40-80 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women evaluated for pelvic mass at one institution **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** None	1) Logistic regression fitted model: Z = -5.39224 + 2.35132 x US + 2.81806 x PE + 1.58268 x CA-125 + 0.607183 x CA-72-4 - 1.11594 x RI Each variable is "suspicious" (1) or "not suspicious" (0) 2) TVUS (with transabdominal imaging if large tumor) Interpreted according to Valentin et al.: unilocular (1); multilocular (2); unilocular solid cyst (3); multilocular solid cyst (4); solid tumor (5) 3) PE physical exam by standard protocol. Examiner was asked to guess benign or malignant. This clinical impression was used in model. No mention of blinding to other data. 4) CA-125 levels > 35 U/ml considered abnormal 5) CA-72-4 levels > 3 U/ml considered abnormal 6) Intratumoral resistance index (RI) was evaluated by color Doppler. RI < 0.4	1) Logistic regression model prediction Dis+ / Dis- / Tot T+: 17 / 5 / 22 T-: 5 / 45 / 50 Tot: 22 / 50 / 72 Value (Lower 95% CI, Upper 95% CI) Se: 77.3% (59.8%, 94.8%) Sp: 90.0% (81.7%, 98.3%) PPV: 77.3% (59.8%, 94.8%) NPV: 90.0% (81.7%, 98.3%) **Results were reported, but have not been abstracted, for the following combinations:** Se and Sp were given for individual components (PE, US, CA-125, CA-72-4) as well as various combinations (not clearly defined how these were operationalized – but apparently not with fitted LR models)	**Comments:** --Exclusion of 18 women with benign exams (not verified) --Authors do not state cutoff for positive in their model **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
				considered abnormal.		
Biagiotti, Desii, Vanzi, et al., 1999 **#2990**	**Geographical location:** Florence, Italy **Dates:** NR **Size of population:** 207 women (226 adnexal masses) **Case series; diagnostic test study** **Reference standard:** Histopathological diagnosis **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp **Blinding:** None **Definition of positive and negative on screening test:** Determined by model fit	**Age:** Range: 21-74 **Menopausal status (n [%]):** Pre (< 45): 146 (70.5%) Post (> 55): 61 (29.5%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** 19 patients had bilateral adnexal masses **Inclusion criteria:** Women undergoing TVUS before surgery for adnexal masses **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Logistic regression models built from five training subsets of the data; candidate variables included: age, and 6 US variables: mean diameter of mass (mm), multilocularity, papillary projections, random echogenicity, peak systolic velocity (cm/sec), RI 2) Artificial neural network using the 5 predictor variables identified from forward stepwise selection in LR training set models: age, and 6 US variables: papillary projections, random echogenicity, peak systolic velocity (cm/sec), RI 3) TVUS with transabdominal US for large masses. Color Doppler imaging used to calculate RI.	1) Logistic regression models T+ indicates > 50% probability of malignancy. Analysis based on number of tumors rather than number of patients. Dis+ / Dis- / Tot T+: 43 / 6 / 49 T-: 8 / 169 / 177 Tot: 51 / 175 / 226 Value (Lower 95% CI, Upper 95% CI) Se: 84.3% (74.3%, 94.3%) Sp: 96.6% (93.9%, 99.3%) PPV: 87.8% (78.6%, 96.9%) NPV: 95.5% (92.4%, 98.5%) 2) Artificial neural networks Analysis based on number of tumors rather than number of patients. Cutoff > 50%. Dis+ / Dis- / Tot T+: 49 / 4 / 53 T-: 2 / 171 / 173 Tot: 51 / 175 / 226 Value (Lower 95% CI, Upper 95% CI) Se: 96.1% (90.8%, 100.0%) Sp: 97.7% (95.5%, 99.9%) PPV: 92.5% (85.3%, 99.6%) NPV: 98.8% (97.3%, 100.0%)	**Comments:** --Analysis based on number of tumors rather than number of patients --This study used almost exclusively US predictors; age was the only non-US predictor. --Formula not given for multiple logistic regression model **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: - Explicit validation method?: + (split sample and bootstrap validation)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Chou, Chang, Yao, et al., 1994 **#10930**	**Geographical location:** Tainan, Taiwan **Dates:** Jan 1991-Feb 1993 **Size of population:** 108 women **Registry** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, PPV, NPV **Blinding:** None **Definition of positive and negative on screening test:** See under "Scoring System." Neither TVUS morphology criteria nor color Doppler US intratumoral vessel criteria were explicitly defined/described.	**Age:** Mean: 38 Range: 11-85 **Menopausal status (n [%]):** Pre (< 45): 84 (78%) Post (> 55): 19 (18%) Premenarchal: 5 (4%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women undergoing surgery for adnexal tumors **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 108 (100%) by CT or US **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) TVUS - morphology 2) Color Doppler US to visualize intratumoral vessel and flow velocity 3) RI with cutpoint of 0.5 4) Serum CA-125 level with cutpoint of 35 U/ml or 65 U/ml	1) RI < 0.5 or CA-125 > 35 U/ml |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 25 | 2 | 27 | | T- | 0 | 81 | 81 | | Tot | 25 | 83 | 108 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 100.0% | 88.0% | 100.0% | | Sp | 97.0% | 93.3% | 100.0% | | PPV | 92.6% | 82.7% | 100.0% | | NPV | 100.0% | 96.3% | 100.0% | 2) RI < 0.5 or CA-125 > 65 U/ml |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 25 | 0 | 25 | | T- | 0 | 83 | 83 | | Tot | 25 | 83 | 108 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 100.0% | 88.0% | 100.0% | | Sp | 100.0% | 96.4% | 100.0% | | PPV | 100.0% | 88.0% | 100.0% | | NPV | 100.0% | 96.4% | 100.0% |	**Comments:** --Poor description of morphologic criteria for positive ultrasound --2x2 tables do not agree with Se, Sp, PPV, and NPV statistics reported; not consistent with rounding error alone **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: - Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Davies, Jacobs, Woolas, et al., 1993 **#4720**	**Geographical location:** London, UK **Dates:** NR **Size of population:** 124 women **Retrospective series** **Reference standard:** Histopathological or operative report diagnosis **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, ROC curves **Blinding:** None **Definition of positive and negative on screening test:** Multiple cutoffs tested; measures defined as per Jacobs, Oram, Fairbanks, et al., 1990 (#6820)	**Age:** NR **Menopausal status (n [%]):** Pre: 86 (69%) Post: 38 (31%) ≥ 1 year amenorrhea or age > 50 if status post hysterectomy **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Consecutive admissions for surgical exploration of adnexal mass **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RMI1 of Jacobs, Oram, Fairbanks, et al., 1990 (#6820) RMI = U x M x CA-125 2) Menopausal status (M) scored as follows: Postmenopausal (> 1 year of amenorrhea) = 3 Premenopausal =1 3) Serum CA-125 level by RIA kit (CIS Bioindustries, France). Scored as absolute value. 4) Pelvic US score (U; transabdominal) based on presence of the following factors: multilocularity (1) solid areas (1) bilaterality (1) ascites (1) extraovarian tumors (1) If total is 0, then U = 0 If total is 1, then U = 1 If total ≥ 2, then U = 3	1) RMI1 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820]) with cutoff value of 200 for T+ Dis+ / Dis- / Tot T+: 33 / 11 / 44 T-: 4 / 76 / 80 Tot: 37 / 87 / 124 Value (95% CI) Se: 89.0% (78.9% - 99.1%) Sp: 87.0% (79.9% - 94.1%) PPV: 75.0% (62.2% - 87.8%) NPV: 95.0% (90.2% - 99.8%) **Results were reported, but have not been abstracted, for the following combinations:** CA-125 with cutoff values of 30, 50, 70, 90, 120 Ultrasound score with cutoff values of 1 and 3 Menopausal status (post vs pre) RMI with cutoffs of 25, 50, 75, 100, 150, 250	**Comments:** --Borderlines counted as malignant **Quality assessment:** Reference standard: - (operative diagnosis accepted in absence of histopathological diagnosis) Verification bias: + Test reliability/variability: +/- (intra- and inter-assay coefficients of variation > 10% for CA-125) Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: - Explicit validation method?: + (this is a validation study of RMI1)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Dowd, Quinn, Rome, et al., 1993 **#4680**	**Geographical location:** Melbourne, Australia **Dates:** NR **Size of population:** 264 women **Registry** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp **Blinding:** None **Definition of positive and negative on screening test:** RMI > 200	**Age:** Mean: 50.2 Range: 15-89 **Menopausal status (n [%]):** Pre: 121 (46%) Post: 143 (54%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Serum CA-125 performed for evaluation of pelvic mass **Exclusion criteria:** No suspected pelvic mass; CA-125 obtained for screening only; no histopathology	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** CA-125 Menopausal status Ultrasound score	1) CA-125 Absolute value (capped at 500) 2) Ultrasound score (U); criteria not described (0, 1, or 3 depending on the presence of particular features) 3) Menopausal score (M) Postmenopausal (> 1 year amenorrhea) = 3 Premenopausal = 1 RMI1 = U x M x CA-125 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820]); cutoff > 2000	1) RMI1 (n = 180 patients) Se = 70% Sp = 94% Can't calculate 2x2 (marginals not reported for this 180-patient subset for whom RMI1 was available). No stratification by age or menopausal status.	**Comments:** --Unable to determine 2x2 table from available information – reported that: "There was sufficient data to use the RMI of Jacobs, Oram, Fairbanks, et al., 1990 (#6820), in 180 patients: the Se was 0.70 and the Sp was 0.94." No singular solution or marginals were reported. --Borderlines counted as malignant **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method? + (this is validation study)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Gadducci, Capriello, Bartolini, et al., 1988 **#6650**	**Geographical location:** Pisa, Italy **Dates:** NR **Size of population:** 119 patients **Diagnostic test study among referral series** **Reference standard:** Histopathology; specimen obtained at laparotomy **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp **Blinding:** No **Definition of positive and negative on screening test:** NA	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Undergoing laparotomy for adnexal mass **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US score based on: Shape (rounded-0 patients; polycyclic – 2 patients; poorly defined – 4 patients) Ascites (absent – 0 patients; present – 4 patients); Outline (regular – 0 patients; poorly defined – 2 patients; thickened – 4 patients) Structure (anechoic or mildly echogenic – 0 patients; homogeneous echogenic – 1 patient; multilocular – 2 patients; anechoic with echogenic areas – 3 patients; echogenic with anechoic areas – 4 patients) Total score determined by summing points for each parameter (range 0-16 patients) T- < 10 T+ ≥ 10 2) CA-125 solid phase sandwich radioimmunoassay (Centodor kit) T+ ≥ 65 U/ml T- < 65 U/ml	1) Combined evaluation of US and CA-125 assay: T+ if either (or both) abnormal; T- if both normal Dis+ / Dis- / Tot T+: 33 / 5 / 38 T-: 2 / 78 / 80 Tot: 35 / 83 / 118 Value / Lower 95% CI / Upper 95% CI Se: 94.3% / 86.6% / 100.0% Sp: 94.0% / 88.9% / 99.1% PPV: 86.8% / 76.1% / 97.6% NPV: 97.5% / 94.1% / 100.0%	**Comments:** --Scoring system was likely an afterthought, not formally developed or well defined **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Guerriero, Ajossa, Risalvato, et al., 1998 **#3400**	**Geographical location:** Cagliari, Italy **Dates:** Jan 1996-May 1997 **Size of population:** 192 adnexal masses in 178 women from among 240 eligible referred women **Registry** Consecutive **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** Yes	**Age:** Mean (SD): 41 (15) Range: 14-77 **Menopausal status (n [%]):** Pre: 127 (71%) Post: 51 (29%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Consecutive women with adnexal mass based on palpation or sonography **Exclusion criteria:** Pregnant	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) B-mode TVUS; Morphological criteria defined for benign disorders 2) Color Doppler imaging with PI and RI calculated. If no Doppler waveforms were obtained, then the result was considered negative. 3) CA-125 (> 35 U/ml or > 65 U/ml)	1) CDE and CA-125 ≥ 65 U/ml (all patients – numbers are masses) |	Dis+	Dis-	Tot
---	---	---	---
T+	22	5	27
T-	11	140	151
Tot	33	145	178 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	66.7%	50.6%	82.8%
Sp	96.6%	93.6%	99.5%
PPV	81.5%	66.8%	96.1%
NPV	92.7%	88.6%	96.9% | 2) CDE and CA-125 ≥ 65 U/ml (postmenopausal women only – numbers are masses) |	Dis+	Dis-	Tot
---	---	---	---
T+	17	0	17
T-	9	27	36
Tot	26	27	53 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	65.4%	47.1%	83.7%
Sp	100.0%	88.9%	100.0%
PPV	100.0%	82.4%	100.0%
NPV	75.0%	60.9%	89.1% | 3) CDE and CA-125 ≥ 65 U/ml (premenopausal women only – numbers are masses) |	Dis+	Dis-	Tot
---	---	---	---
T+	5	5	10
T-	2	113	115
Tot	7	118	125 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	71.4%	38.0%	100.0%
Sp	95.8%	92.1%	99.4%		**Comments:** --Study described intra-observer variation for CDE, RI, and PI **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +/- Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

PPV 50.0% 19.0% 81.0%

4) CDE and CA-125 ≥ 35 U/ml (postmenopausal women only – numbers are masses)

	Dis+	Dis-	Tot
T+	21	1	22
T-	5	26	31
Tot	26	27	53

	Value	Lower 95% CI	Upper 95% CI
Se	80.8%	65.6%	95.9%
Sp	96.3%	89.2%	100.0%
PPV	95.5%	86.8%	100.0%
NPV	83.9%	70.9%	96.8%

5) B-mode and CA-125 ≥ 65 U/ml (postmenopausal women only – numbers are masses)

	Dis+	Dis-	Tot
T+	17	0	17
T-	9	27	36
Tot	26	27	53

	Value	Lower 95% CI	Upper 95% CI
Se	65.4%	47.1%	83.7%
Sp	100.0%	88.9%	100.0%
PPV	100.0%	82.4%	100.0%
NPV	75.0%	60.9%	89.1%

6) B-mode and CA-125 ≥ 35 U/ml (postmenopausal women only – numbers are masses)

	Dis+	Dis-	Tot
T+	21	1	22
T-	5	26	31
Tot	26	27	53

	Lower	Upper

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results				Comments/Quality Scoring
						Value	95% CI	95% CI	
					Se	80.8%	65.6%	95.9%	
					Sp	96.3%	89.2%	100.0%	
					PPV	95.5%	86.8%	100.0%	
					NPV	83.9%	70.9%	96.8%	

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Guerriero, Alcazar, Coccia, et al., 2002 **#2130**	**Geographical location:** Cagliari, Florence, and Navarra, Italy **Dates:** Apr 1997-Jul 2000 **Size of population:** 826 masses in 789 women from a potential study population of 1020 women **Registry** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, PPV, NPV **Blinding:** None **Definition of positive and negative on screening test:** Yes	**Age:** Mean (SD): 40 (14) Range: 14-81 **Menopausal status (n [%]):** Pre: 617 (78%) Post: 172 (22%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** All women scheduled for surgery for the presence of a persistent adnexal mass **Exclusion criteria:** Anechoic unilocular or bilocular cystic mass with a thin regular wall without endocystic vegetation (n = 234)	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US morphology 2) Color Doppler US 3) CA-125	1) B-mode and CA-125 > 35 U/ml (postmenopausal women only – numbers are masses) 2) Color Doppler and CA-125 > 35 U/ml (postmenopausal women only – numbers are masses) 3) B-mode and CA-125 > 35 U/ml (all women – numbers are masses)	**Comments:** --19 of 132 ovarian carcinomas were low-malignancy potential tumors; 6 masses were metastases from non-ovarian primaries **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

1) B-mode and CA-125 > 35 U/ml (postmenopausal women only – numbers are masses)

	Dis+	Dis-	Tot
T+	73	6	79
T-	20	74	94
Tot	93	80	173

	Value	Lower 95% CI	Upper 95% CI
Se	78.5%	70.1%	86.8%
Sp	92.5%	86.7%	98.3%
PPV	92.4%	86.6%	98.2%
NPV	78.7%	70.4%	87.0%

2) Color Doppler and CA-125 > 35 U/ml (postmenopausal women only – numbers are masses)

	Dis+	Dis-	Tot
T+	71	3	74
T-	22	76	98
Tot	93	79	172

	Value	Lower 95% CI	Upper 95% CI
Se	76.3%	67.7%	85.0%
Sp	96.2%	92.0%	100.0%
PPV	95.9%	91.5%	100.0%
NPV	77.6%	69.3%	85.8%

3) B-mode and CA-125 > 35 U/ml (all women – numbers are masses)

	Dis+	Dis-	Tot
T+	113	30	143
T-	34	649	683
Tot	147	679	826

	Value	Lower 95% CI	Upper 95% CI
Se	76.9%	70.1%	83.7%
Sp	95.6%	94.0%	97.1%

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

PPV 79.0% 72.3% 85.7%
NPV 95.0% 93.4% 96.7%

4) Color Doppler and CA-125 > 35 U/ml (all women – numbers are masses)

	Dis+	Dis-	Tot
T+	112	10	122
T-	35	669	704
Tot	147	679	826

	Value	Lower 95% CI	Upper 95% CI
Se	76.2%	69.3%	83.1%
Sp	98.5%	97.6%	99.4%
PPV	91.8%	86.9%	96.7%
NPV	95.0%	93.4%	96.6%

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Jacobs, Oram, Fairbanks, et al., 1990 **#6820**	**Geographical location:** London, UK **Dates:** NR **Size of population:** 143 women **Unclear whether retrospective or prospective series; diagnostic test study** **Reference standard:** Histopathological or operative report diagnosis **Reference standard applied to all test negatives?:** No **Statistical tests used:** Se, Sp, ROC curves, LR **Blinding:** Clinical assessment blind only to CA-125 level **Definition of positive and negative on screening test:** No a priori cutoff	**Age:** Mean: 51.8 **Menopausal status (n [%]):** Pre: 61 Post: 82 > 1 year of amenorrhea or age > 50 if status post hysterectomy **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Admitted for elective surgical investigation of an adnexal mass **Exclusion criteria:** None	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	RMI1 – clinical prediction rule developed from logistic regression model based on ultrasound, CA-125 and menopausal status RMI1 = U x M x CA-125 1) Ultrasound score (U; transabdominal) – 1 point for each of the following: multilocular cyst, evidence of solid areas, evidence of metastases, presence of ascites, and bilateral lesions U = 0 for US score of 0 =1 for US score of 1 = 3 for US score ≥ 2 3) Serum CA-125 by RIA (Abbott Labs, Chicago) 4) Menopausal status (M): 1 if premenopausal 3 if postmenopausal	1) RMI1 with cutoff value of 200 |	Dis+	Dis-	Tot
---	---	---	---
T+	35	3	38
T-	6	95	101
Tot	41	98	139 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	85.4%	74.5%	96.2%
Sp	96.9%	93.5%	100.0%
PPV	92.1%	83.5%	100.0%
NPV	94.1%	89.4%	98.7% | 2) RMI1 with cutoff value of 250 (maximum specificity) |	Dis+	Dis-	Tot
---	---	---	---
T+	32	1	33
T-	9	97	106
Tot	41	98	139 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	78.0%	65.3%	90.7%
Sp	99.0%	97.0%	100.0%
PPV	97.0%	91.1%	100.0%
NPV	91.5%	86.2%	96.8% | 3) RMI1 with cutoff value of 25 (maximum Se) |	Dis+	Dis-	Tot
---	---	---	---
T+	41	37	78
T-	0	61	61
Tot	41	98	139 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	92.7%	100.0%
Sp	62.2%	52.6%	71.8%
PPV	52.6%	41.5%	63.6%
NPV	100.0%	95.1%	100.0%		**Comments:** --Small data set relative to number of predictors examined; no validation; no a priori definition of positive items --Borderlines counted as malignant **Quality assessment:** Reference standard: - (when no specimen was sent for histopathology, the surgical diagnosis was assumed to be correct) Verification bias: + Test reliability/variability: +/- (coefficient of variation < 10% for CA-125) Sample size: - Statistical tests: + Blinding: - (CA-125 only) Definition of +/- on screening test: - Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
					Results were reported, but have not been abstracted, for the following combinations: RMI 1with cutoff values of 50, 75, 100, 150 Se / Sp 50: 95.1% / 76.5% 75: 92.7% / 84.7% 100: 85.4% / 87.8% 150: 85.4% / 93.9%	
Lu, Van Gestel, Suykens, et al., 2003 **#1730**	**Geographical location:** Leuven, Belgium **Dates:** 1994-1999 1994-1997 (training) 1997-1999 (test) **Size of population:** 525 women 265 (training set) 160 (test set) **Retrospective case series** **Reference standard:** Histopathologic diagnosis **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, ROC curves **Blinding:** None	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Women referred for US to single ultrasonographer with persistent extrauterine pelvic mass which was subsequently surgically removed **Exclusion criteria:** No preoperative CA-125 assay (n = 100)	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 525 (100%) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Multivariable models developed using least squares support vector machine (LS-SVM) classifiers in a Bayesian evidence framework. Model built on 265 patient training set; tested on 160-patient test set. Candidate variables included 27 demographic, serum marker, color Doppler, B-mode US, US morphologic, and US echogenicity variables. Variables were chosen using forward selection. Six different models were built and tested 2) LR1- logistic regression 3) LS-SVM1 (Lin) 4) LS-SVM1 (RBF) 5) LR2	1) RMI1 with cutoff value of 75 Dis+ / Dis- / Tot T+: 44 / 27 / 71 T-: 10 / 79 / 89 Tot: 54 / 106 / 160 Value (Lower 95% CI – Upper 95% CI) Se: 81.5% (71.1% – 91.8%) Sp: 74.5% (66.2% – 82.8%) PPV: 62.0% (50.7% – 73.3%) NPV: 88.8% (82.2% – 95.3%) AUC = 0.8733 (± 0.0298 SE) 2) RMI1 with cutoff value of 100 Dis+ / Dis- / Tot T+: 40 / 21 / 61 T-: 14 / 85 / 99 Tot: 54 / 106 / 160 Value (Lower 95% CI – Upper 95% CI) Se: 74.1% (62.4% – 85.8%) Sp: 80.2% (72.6% – 87.8%) PPV: 65.6% (53.7% – 77.5%) NPV: 85.9% (79.0% – 92.7%) **Results were reported, but have not been abstracted, for the following**	**Comments:** --Data from Table 2 (page 296) --Models 2-7 parameters not specified --This paper written to demonstrate feasibility of new modeling approach and application to ovarian mass; does not provide model for use in clinical decisionmaking **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: - Explicit validation method?: +

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
	Definition of positive and negative on screening test:			6) LS-SVM2 (Lin) 7) LS-SVM2 (RBF) 8) RMI1 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820])	**combinations:** Models 2-7 have Se, Sp, PPV and NPV for 3 to 4 cutoff values each and AUC reported. Each performs better than RMI1.	

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Ma, Shen, and Lang, 2003 **#1900**	**Geographical location:** Beijing, China **Dates:** Jan 1998-Jun 1999 **Size of population:** 140 women **Single-institution retrospective case series** **Reference standard:** Not stated explicitly, but implied that all patients had pathological classification **Reference standard applied to all test negatives?:** Yes, presumably **Statistical tests used:** Se, Sp, PPV **Blinding:** None **Definition of positive and negative on screening test:** Used definitions of Tingulstad, Hagen, Skjeldestad, et al., 1996 (#3890), but analyzed multiple cutoff values	**Age:** Range: 30 - NR **Menopausal status (n [%]):** Pre: 89 (64%) Post: 51 (36%) > 1 year of amenorrhea or age > 50 if status post hysterectomy **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** "Ovarian neoplasm" patients over 30 years admitted to a single institution **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) RMI2 (Tingulstad, Hagen, Skjeldestad, et al., 1996 [#3890]) – clinical prediction rule developed from logistic regression model based on ultrasound (U), CA-125, and menopausal status (M) RMI2 = U x M x CA-125 2) Ultrasound (U; transabdominal) scored 1 point for each of the following characteristics: multilocular cyst, evidence of solid areas, evidence of metastases, presence of ascites, and bilateral lesions. U = 1 for total score of 0-1 4 for total score ≥ 2 3) Serum CA-125 4) Menopausal status (M): 1 if premenopausal 4 if postmenopausal	1) RMI2 cutoff value 200 Dis+ / Dis- / Tot T+: 55 / 12 / 67 T-: 8 / 65 / 73 Tot: 63 / 77 / 140 Value (Lower 95% CI – Upper 95% CI) Se: 87.3% (79.1% – 95.5%) Sp: 84.4% (76.3% – 92.5%) PPV: 82.1% (72.9% – 91.3%) NPV: 89.0% (81.9% – 96.2%) 2) RMI2 cutoff value 100 Dis+ / Dis- / Tot T+: 59 / 21 / 80 T-: 4 / 56 / 60 Tot: 63 / 77 / 140 Value (Lower 95% CI – Upper 95% CI) Se: 93.7% (87.7% – 99.7%) Sp: 72.3% (62.3% – 82.3%) PPV: 73.8% (64.1% – 83.4%) NPV: 93.3% (87.0% – 99.6%) **Results were reported, but have not been abstracted, for the following combinations:** RMI2 with cutoff of 50, 400, 1000	**Comments:** --Relatively poor quality of reporting **Quality assessment:** Reference standard: +/- Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: - Explicit validation method?: + (this is a validation study of RMI2)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Maggino, Gadducci, D'Addario, et al., 1994 **#4500**	**Geographical location:** Italy **Dates:** Mar 1991-Mar 1992 **Size of population:** 383 women 48 excluded for inadequate data 335 evaluable 45 benign cysts 290 surgical cases **Case series** **Reference standard:** Histopathological diagnosis in 290 women; clinical and US followup in 45 patients with benign-appearing cysts **Reference standard applied to all test negatives?:** 45 women with US findings indicating benign cyst and CA-125 < 35 U/ml not verified **Statistical tests used:** Se, Sp, PPV, NPV **Blinding:** None **Definition of positive and negative on screening test:** Yes	**Age:** Mean: 61.9 Range: 40-91 **Menopausal status (n [%]):** Post: 290 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Postmenopausal women with pelvic mass (intra- or extra-adnexal) **Exclusion criteria:** Premenopause; previous malignant neoplasia (except breast cancer); previous bilateral adnexectomy; if > 55 years of age, previous hysterectomy for non-tumoral disease	**Symptomatic (n [%]):** 209 (72%) **Detected by exam (n [%]):** All patients **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Gynecologic examination 2) US (transabdominal) Classified as: Benign if size < 5 cm; wall clear, thin; non-echogenous content; no septa or ≤ 3 thin septa; no free liquid in pouch of Douglas; Equivocal if size 5-10 cm; wall clear, smooth, thick; hypoechogenic liquid content or solild homogeneous content; >3 thin septa; thick but regular septa; absence of endocystic vegetation; absence of free peritoneal fluid; Malignant if none of above features observed 3) CA-125: T- < 35 U/ml Borderline 35-65 U/ml T+ > 65 U/ml 4) Combination of US and CA-125 T+ if either individual test abnormal T- if neither test abnormal	1) Combined US (Gr III possibly malignant) or CA-125 (> 65 U/ml) as T+; else, T- Analysis limited to ovarian tumors Dis+ / Dis- / Tot T+: 100 / 36 / 136 T-: 6 / 98 / 104 Tot: 106 / 134 / 240 Value (Lower 95% CI – Upper 95% CI) Se: 94.3% (89.9% – 98.7%) Sp: 73.1% (65.6% – 80.6%) PPV: 73.5% (66.1% – 80.9%) NPV: 94.2% (89.7% – 98.7%) 2) Combined US (Gr III possibly malignant) or CA-125 (> 35 U/ml) as T+; else, T- Analysis limited to ovarian tumors Dis+ / Dis- / Tot T+: 102 / 45 / 147 T-: 4 / 89 / 93 Tot: 106 / 134 / 240 Value (Lower 95% CI – Upper 95% CI) Se: 96.2% (92.6% – 99.9%) Sp: 66.4% (58.4% – 74.4%) PPV: 69.4% (61.9% – 76.8%) NPV: 95.7% (91.6% – 99.8%) 3) Combined US (Gr II-III possibly malignant or borderline) or CA-125 (> 65 U/ml) as T+; else, T- Analysis limited to ovarian tumors Dis+ / Dis- / Tot T+: 106 / 77 / 183 T-: 0 / 57 / 57 Tot: 106 / 134 / 240	**Comments:** --Transabdominal US only; no endovaginal scanning done --Note typographical errors in Table 10: Se 93.34 should be 94.34 and 65.70 should be 95.70 --Limiting analysis to cases in which both tests (CA-125 and US) are concordant is not clinically useful, and Se, Sp, PPV, and NPV from this table should be ignored **Quality assessment:** Reference standard: + Verification bias: - (45 women not verified, but were followup with US and clinical exam) Test reliability/variability: - Sample size: - Statistical tests: - (misguided statistical analysis) Blinding: - Definition of +/- on screening test: - Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	97.2%	100.0%
Sp	42.5%	34.2%	50.9%
PPV	57.9%	50.8%	65.1%
NPV	100.0%	94.7%	100.0%

4) Combined US (Gr II-III possibly malignant or borderline) or CA-125 (>35 U/ml) as T+; else, T-
Analysis limited to ovarian tumors

	Dis+	Dis-	Tot
T+	106	82	188
T-	0	52	52
Tot	106	134	240

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	97.2%	100.0%
Sp	38.8%	30.6%	47.1%
PPV	56.4%	49.3%	63.5%
NPV	100.0%	94.2%	100.0%

5) CA-125 > 65

	Dis+	Dis-	Tot
T+	76	10	86
T-	30	124	154
Tot	106	134	240

	Value	Lower 95% CI	Upper 95% CI
Se	71.7%	63.1%	80.3%
Sp	92.5%	88.1%	97.0%
PPV	88.4%	81.6%	95.1%
NPV	80.5%	74.3%	86.8%

6) US
T+ Class III (possibly malignant)
analysis limited to ovarian tumors (non-ovarian tumors excluded)

	Dis+	Dis-	Tot
T+	90	30	120

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

T-	16	104	120
Tot	106	134	240

	Value	Lower 95% CI	Upper 95% CI
Se	84.9%	78.1%	91.7%
Sp	77.6%	70.6%	84.7%
PPV	75.0%	67.3%	82.7%
NPV	86.7%	80.6%	92.7%

Results were reported, but have not been abstracted, for the following combinations:
No other combinations, but the following individual tests:
US (Class II-II possibly malignant or borderline as positive)
CA-125 (> 35 U/ml as positive)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Mancuso, De Vivo, Triolo, et al., 2004 **#1610**	**Geographical location:** Messina, Italy **Dates:** NR **Size of population:** 125 women **Case series** **Reference standard:** Histopathological diagnosis **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, PPV, NPV **Blinding:** None **Definition of positive and negative on screening test:** CA-125 - >35 U/ml US, defined	**Age:** Mean: 42.2 Range: 18-82 **Menopausal status (n [%]):** Pre: 76 (61%) Post: 49 (39%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** Nulliparous: 62 (50%) Multiparous: 63 (50%) **Inclusion criteria:** Women admitted for adnexal mass **Exclusion criteria:** None	**Symptomatic (n [%]):** Pain 68 (54%) Menstrual disorder 22 (18%) Urinary/intestinal 5 (4%) Asymptomatic 30 (24%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) US (transvaginal) Positive if: solid structure or cystic but complex; irregular walls, endocystic vegetations, thick septa; bilateral lesions; ascites 2) CA-125 positive if > 35 U/ml 3) Combinations of US, age (> 50 years), CA-125, menopause (post)	1) Combination of US + age > 50 |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 13 | 10 | 23 | | T- | 1 | 101 | 102 | | Tot | 14 | 111 | 125 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | 92.9% | 79.4% | 100.0% | | Sp | 91.0% | 85.7% | 96.3% | | PPV | 56.5% | 36.3% | 76.8% | | NPV | 99.0% | 97.1% | 100.0% | 2) Combination of CA-125 + age > 50 |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 13 | 4 | 17 | | T- | 1 | 107 | 108 | | Tot | **14** | **111** | 125 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | **91.0%** | 76.0% | 100.0% | | Sp | **96.6%** | 93.2% | 100.0% | | PPV | 76.5% | 56.3% | 96.6% | | NPV | 99.1% | 97.3% | 100.0% | 3) Combination of US + menopause |	Dis+	Dis-	Tot | |---|---|---|---| | T+ | 14 | 13 | 27 | | T- | 0 | 98 | 98 | | Tot | **14** | **111** | 125 | |	Value	Lower 95% CI	Upper 95% CI | |---|---|---|---| | Se | **100.0%** | 78.6% | 100.0% | | Sp | **88.7%** | 82.8% | 94.6% | | PPV | 51.9% | 33.0% | 70.7% | | NPV | 100.0% | 96.9% | 100.0% | 4) Combination of CA-125 + menopause	**Comments:** --2x2 cell numbers do not result in exact figures for PPV as reported in Table 4 **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	14	4	18
T-	0	107	107
Tot	**14**	**111**	125

	Value	Lower 95% CI	Upper 95% CI
Se	**100.0%**	78.6%	100.0%
Sp	**96.6%**	93.2%	100.0%
PPV	77.8%	58.6%	97.0%
NPV	100.0%	97.2%	100.0%

5) Combination of CA-125 + US

	Dis+	Dis-	Tot
T+	14	6	20
T-	0	105	105
Tot	**14**	**111**	125

	Value	Lower 95% CI	Upper 95% CI
Se	**100.0%**	78.6%	100.0%
Sp	**94.4%**	90.1%	98.7%
PPV	70.0%	49.9%	90.1%
NPV	100.0%	97.1%	100.0%

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Manjunath, Pratap-kumar, Sujatha, et al., 2001 **#2510**	**Geographical location:** Manipal, India **Dates:** Jan 97-Aug 99 **Size of population:** 152 **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes – applied to all in population regardless of RMI score **Statistical tests used:** Chi-square, Se, Sp, PPV, NPV **Blinding:** NR **Definition of positive and negative on screening test:** Ultrasound: 1 point each for: multilocular, solid areas, ascites, bilateral, and intra-abdominal metastases Menopausal status: > 1 yr amenorrhea or age > 50	**Age:** NR **Menopausal status (n [%]):** Pre: 84 (57%) Post: 64 (43%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients who underwent surgery for a pelvic mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Ultrasound score (U) 1 point each for: multilocular, solid areas, ascites, bilateral, and intra-abdominal metastases 2) CA-125 3) Menopausal status 4) RMI1 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820]) RMI1 = U x M x CA-125, where M = 1 or 3 (pre or post) and U = 0 (no points), 1 (one point) or 3 (2 or more points on ultrasound) 5) RMI2 (Tingulstad, Hagen, Skjeldestad, et al., 1996 [#3890]) RMI2 = U x M x CA-125, where M = 1 or 4 (pre or post) and U = 1 (score 0 or 1) or 4 (score 2 or more) 6) RMI3 (Tingulstad et al 1999) RMI3 = U x M x CA-125, where M = 1 or 3 (pre or post), U = 1 (score 0 or 1) or 3 (score at least 2)	Note: Borderline tumors not counted, so n = 148 (appropriate) Note: Results are given for 11 different cutoff scores. The authors recommend cutoff of 200 for all 3 RMIs to minimize false positives. 1) RMI1 (cutoff 200) Dis+ / Dis- / Tot T+: 68 / 5 / 73 T-: 25 / 50 / 75 Tot: 93 / 55 / 148 Value / Lower 95% CI / Upper 95% CI Se: 73.0% / 64.0% / 82.0% Sp: 91.0% / 83.4% / 98.6% PPV: 93.2% / 87.4% / 98.9% NPV: 66.7% / 56.0% / 77.3% 2) RMI2 (cutoff 200) Dis+ / Dis- / Tot T+: 71 / 10 / 81 T-: 22 / 45 / 67 Tot: 93 / 55 / 148 Value / Lower 95% CI / Upper 95% CI Se: 76.0% / 67.3% / 84.7% Sp: 82.0% / 71.8% / 92.2% PPV: 87.7% / 80.5% / 94.8% NPV: 67.2% / 55.9% / 78.4% 3) RMI3 (cutoff 125) Dis+ / Dis- / Tot T+: 74 / 11 / 85 T-: 19 / 44 / 63 Tot: 93 / 55 / 148 Lower / Upper	**Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: +/- (no a priori cutoff) Explicit validation method?: + (validation of prior reports on the RMI)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

	Value	95% CI	95% CI
Se	**80.0%**	71.9%	88.1%
Sp	**80.0%**	69.4%	90.6%
PPV	87.1%	79.9%	94.2%
NPV	69.8%	58.5%	81.2%

4) RMI3 (cutoff 200)

	Dis+	Dis-	Tot
T+	69	5	74
T-	24	50	74
Tot	**93**	**55**	148

	Value	Lower 95% CI	Upper 95% CI
Se	**74.0%**	65.1%	82.9%
Sp	**91.0%**	83.4%	98.6%
PPV	93.2%	87.5%	99.0%
NPV	67.6%	56.9%	78.2%

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Mol, Boll, De Kanter, et al., 2001 **#5780**	**Geographical location:** Utrecht, the Netherlands **Dates:** 1991-1998 **Size of population:** 170 women **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** ROC curves **Blinding:** NR **Definition of positive and negative on screening test:** Per the original reports of the models (cutoffs not specified here)	**Age:** Mean: 46.6 Range: 20-89 **Menopausal status (n [%]):** Pre: 109 (64%) Post: 61 (36%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Surgery for an adnexal mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Tailor 1997 ref 19 US, Doppler, age, AUC 0.81 2) Prompeler 1997 ref 22, AUC 0.73 US, menopausal status 3) Jacobs 1990 ref 13 menopause, US, CA-125, AUC 0.83 4) Jacobs 1993 ref 23 menopause, US CA-125, AUC 0.86 5)Tingulstad 1996 ref 12 menopause, US, CA-125, AUC 0.83 6) Timmerman 1999 ref 9 menopause, US, Doppler, CA-125 (neural network), AUC 0.84 7) Timmerman 1999 ref 9 Menopause, US, CA-125 (neural network), AUC 0.85 8) Timmerman 1999 ref 9 menopause, US, Doppler, CA-125 (logistic regression), AUC 0.85	1) Tailor et al., 1997 model sonography, color Doppler, age Dis+ / Dis- / Tot T+: 27 / 77 / 104 T-: 3 / 63 / 66 Tot: 30 / 140 / 170 Value (Lower 95% CI – Upper 95% CI) Se: 90.0% (79.3% – 100.0%) Sp: 45.0% (36.8% – 53.2%) PPV: 26.0% (17.5% – 34.4%) NPV: 95.5% (90.4% – 100.0%) 2) Prompeler 1997 ref 22, AUC 0.73 U/S, menopausal status Dis+ / Dis- / Tot T+: 27 / 91 / 118 T-: 3 / 49 / 52 Tot: 30 / 140 / 170 Value (Lower 95% CI – Upper 95% CI) Se: 90.0% (79.3% – 100.0%) Sp: 35.0% (27.1% – 42.9%) PPV: 22.9% (15.3% – 30.5%) NPV: 94.2% (87.9% – 100.0%) 3) Jacobs 1990 ref 13 menopause, U/S, CA-125, AUC 0.83 Dis+ / Dis- / Tot T+: 27 / 69 / 96 T-: 3 / 71 / 74 Tot: 30 / 140 / 170 Value (Lower 95% CI – Upper 95% CI) Se: 90.0% (79.3% – 100.0%) Sp: 51.0% (42.7% – 59.3%) PPV: 28.1% (19.1% – 37.1%) NPV: 95.9% (91.5% – 100.0%)	**Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - (this study is testing the models' reliability) Sample size: + Statistical tests: + Blinding: - (NR) Definition of +/- on screening test: + Explicit validation method?: + (this is a validation study)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

4) Jacobs 1993 ref 23
menopause, U/S CA-125, AUC 0.86

	Dis+	Dis-	Tot
T+	27	55	82
T-	3	85	88
Tot	30	140	170

	Value	Lower 95% CI	Upper 95% CI
Se	90.0%	79.3%	100.0%
Sp	61.0%	52.9%	69.1%
PPV	32.9%	22.8%	43.1%
NPV	96.6%	92.8%	100.0%

5) Tingulstad 1996 ref 12
menopause, U/S, CA-125, AUC 0.83

	Dis+	Dis-	Tot
T+	27	69	96
T-	3	71	74
Tot	30	140	170

	Value	Lower 95% CI	Upper 95% CI
Se	90.0%	79.3%	100.0%
Sp	51.0%	42.7%	59.3%
PPV	28.1%	19.1%	37.1%
NPV	95.9%	91.5%	100.0%

6) Timmerman 1999 ref 9
menopause, U/S, Doppler, CA-125 (neural network), AUC 0.84

	Dis+	Dis-	Tot
T+	27	56	83
T-	3	84	87
Tot	30	140	170

	Value	Lower 95% CI	Upper 95% CI
Se	90.0%	79.3%	100.0%
Sp	60.0%	51.9%	68.1%
PPV	32.5%	22.5%	42.6%

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

NPV 96.6% 92.7% 100.0%

7) Timmerman 1999 ref 9
Menopause, U/S, CA-125 (neural network), AUC 0.85

	Dis+	Dis-	Tot
T+	27	76	103
T-	3	64	67
Tot	**30**	**140**	170

	Value	Lower 95% CI	Upper 95% CI
Se	**90.0%**	79.3%	100.0%
Sp	**46.0%**	37.7%	54.3%
PPV	26.2%	17.7%	34.7%
NPV	95.5%	90.6%	100.0%

8) Timmerman 1999 ref 9
menopause, U/S, Doppler, CA-125 (logistic regression), AUC 0.85

	Dis+	Dis-	Tot
T+	27	62	89
T-	3	78	81
Tot	**30**	**140**	170

	Value	Lower 95% CI	Upper 95% CI
Se	**90.0%**	79.3%	100.0%
Sp	**56.0%**	47.8%	64.2%
PPV	30.3%	20.8%	39.9%
NPV	96.3%	92.2%	100.0%

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Morgante, la Marca, Ditto, et al., 1999 **#2900**	**Geographical location:** Siena, Italy **Dates:** Jan 1995-Dec 1997 **Size of population:** 124 women **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Sn,Sp, PPV, NPV, chi-square, ROC curves, Mann-Whitney U, McNemar's **Blinding:** NR **Definition of positive and negative on screening test:** Different cutoffs are reported for RMI1 and RMI2	**Age:** Mean NR Age reported as categorical variable **Menopausal status (n [%]):** Pre: 69 (56%) Post: 55 (44%) (amenorrhea > 1 yr or age > 50 if status post hysterectomy) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age > 30, admitted for surgical evaluation of ovarian mass **Exclusion criteria:** Age < 30, no surgery	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) CA-125 2) Menopausal status (M): age 50 if prior hysterectomy or > 1 yr amenorrhea 3) Ultrasound (U; 1 point each for multilocular, solid, bilateral, ascites, intra-abdominal mets) RMI1 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820]) RMI1 = U x M x CA-125, where U = 0 if score = 0, U = 1 if score = 1, U = 3 if score at least 2; M = 1 if premenopausal, M = 3 if postmenopausal RMI2 (Tingulstad, Hagen, Skjeldestad, et al., 1996 [#3890]) RMI2 = U x M x CA-125, where U = 1 if score = 0 or 1, U = 4 if score at least 2; M = 1 if premenopausal, M = 4 if postmenopausal	1) RMI1 (cutoff 80): Dis+ / Dis- / Tot T+: 25 / 19 / 44 T-: 6 / 74 / 80 Tot: 31 / 93 / 124 Value / Lower 95% CI / Upper 95% CI Se: 81.0% / 67.2% / 94.8% Sp: 80.0% / 71.9% / 88.1% PPV: 56.8% / 42.2% / 71.5% NPV: 92.5% / 86.7% / 98.3% 2) RMI1 (cutoff 200): Dis+ / Dis- / Tot T+: 18 / 5 / 23 T-: 13 / 88 / 101 Tot: 31 / 93 / 124 Value / Lower 95% CI / Upper 95% CI Se: 58.0% / 40.6% / 75.4% Sp: 95.0% / 90.6% / 99.4% PPV: 78.3% / 61.4% / 95.1% NPV: 87.1% / 80.6% / 93.7% 3) RMI2 (cutoff 125): Dis+ / Dis- / Tot T+: 25 / 9 / 34 T-: 6 / 84 / 90 Tot: 31 / 93 / 124 Value / Lower 95% CI / Upper 95% CI Se: 81.0% / 67.2% / 94.8% Sp: 90.0% / 83.9% / 96.1% PPV: 73.5% / 58.7% / 88.4% NPV: 93.3% / 88.2% / 98.5% 4) RMI2 (cutoff 200):	**Comments:** --2 borderline tumors were treated as malignant, which is not exactly correct and may alter results --RMI2 outperformed RMI1 at all cutoff values between 80 and 250 **Quality assessment:** Reference standard: + Verification bias: - (potentially there are patients who did not have surgery whose US findings were less concerning, and we don't have any knowledge of their pathology) Test reliability/variability: + Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: + (tested different cutoffs) Explicit validation method?: + (this is a validation study)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	23	7	30
T-	8	86	94
Tot	31	93	124

	Value	Lower 95% CI	Upper 95% CI
Se	74.0%	58.6%	89.4%
Sp	93.0%	87.8%	98.2%
PPV	76.7%	61.5%	91.8%
NPV	91.5%	85.8%	97.1%

Results were reported, but have not been abstracted, for the following combinations:
RMI1 and RMI2 at following cutoffs:
25, 50, 80, 100, 125, 150, 200, 250

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Obeidat, Amarin, Latimer, et al., 2004 **#1500**	**Geographical location:** Cambridge, England **Dates:** Jan 2000-Dec 2001 **Size of population:** 100 **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Chi-square, Mann-Whitney U **Blinding:** NR **Definition of positive and negative on screening test:** RMI of Jacobs, Oram, Fairbanks, et al., 1990 (#6820) - 4 cutoffs tested: 100, 150, 200, 250	**Age:** Range: 30-NR **Menopausal status (n [%]):** Pre: 27 (27%) Post: 73 (73%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** 100 consecutive women with a pelvic mass who were admitted for laparotomy; age > 30 **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** 100 (100%) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Menopausal status (M = 1 for pre, M = 3 for post); menopause defined as 1 year amenorrhea, or 50 years old for patients with prior hysterectomy 2) Ultrasound score (U): 1 point for each: multilocular, solid, bilateral, ascites, intra-abdominal metastases; U = 0 for score 0, U = 1 for score 1, U = 3 for score ≥ 2 3) CA-125 RMI1 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820]) RMI1 = M x U x CA-125	1) RMI1, cutoff 100: Dis+ / Dis- / Tot T+: 69 / 17 / 86 T-: 3 / 11 / 14 Tot: 72 / 28 / 100 Value / Lower 95% CI / Upper 95% CI Se: 96.0% / 91.5% / 100% Sp: 39.0% / 20.9% / 57.1% PPV: 80.2% / 71.8% / 88.6% NPV: 78.6% / 57.1% / 100% 2) RMI1, cutoff 150: Dis+ / Dis- / Tot T+: 68 / 9 / 77 T-: 4 / 19 / 23 Tot: 72 / 28 / 100 Value / Lower 95% CI / Upper 95% CI Se: 95.0% / 90.0% / 100.0% Sp: 68.0% / 50.7% / 85.3% PPV: 88.3% / 81.1% / 95.5% NPV: 82.6% / 67.1% / 98.1% 3) RMI1, cutoff 200: Dis+ / Dis- / Tot T+: 65 / 3 / 68 T-: 7 / 25 / 32 Tot: 72 / 28 / 100 Value / Lower 95% CI / Upper 95% CI Se: 90.0% / 83.1% / 96.9% Sp: 89.0% / 77.4% / 100% PPV: 95.6% / 90.7% / 100% NPV: 78.1% / 63.8% / 92.4% 4) RMI1 cutoff 250:	**Comments:** --16 borderline tumors were counted as malignant – not strictly true as these are less aggressive tumors which are not treated the same. May bias results. --72% had malignancy. This is a very high prevalence compared to the population that usually presents with a pelvic mass for surgery (usually would be < 10%). Think this may falsely elevate the estimated PPV of the test. **Quality assessment:** Reference standard: + Verification bias: - (possible bias since only patients who had surgery are reported here) Test reliability/variability: + Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: + (several cutoffs tested) Explicit validation method?: + (this is a validation study)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	62	3	65
T-	10	25	35
Tot	72	28	100

	Value	Lower 95% CI	Upper 95% CI
Se	86.0%	78.0%	94.0%
Sp	89.0%	77.4%	100.0%
PPV	95.4%	90.3%	100.0%
NPV	71.4%	56.5%	86.4%

Results were reported, but have not been abstracted, for the following combinations:
"AUC = 0.91"

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Roman, Muder-spach, Stein, et al., 1997 **#6160**	**Geographical location:** Los Angeles, CA **Dates:** Jul 1992-Mar 1994 **Size of population:** 226 women (nonconsecutive) **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** **Statistical tests used:** Chi-square, logistic regression **Blinding:** NR **Definition of positive and negative on screening test:** See Scoring column	**Age:** NR **Menopausal status (n [%]):** Pre: 181 (80%) Post: 45 (20%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Operative intervention for presumed adnexal mass **Exclusion criteria:** Emergency surgery, clinical or radiographic evidence of metastatic disease	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Pelvic exam; fixed, irregular or associated with ascites = suspicious 2) CA-125 > 35 U/ml 3) AFP > 10 ng/ml in non-pregnant patient 4) LDH > 350 U/l 5) TVUS; suspicious masses are those which are cystic with one large or multiple nodules or cystic/solid, and solid masses not arising from the uterus 6) Doppler; pulsatility index less than 1.0 or resistance index ≤ 0.4 are suspicious 7) HCG >15 mIU/ml in a non-pregnant patient	1) Pelvic exam, serum tumor marker, or US positive; postmenopausal women only; low malignant potential tumors counted as Disease + |	Dis+	Dis-	Tot
---	---	---	---
T+	15	15	30
T-	0	13	13
Tot	15	28	43 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	100.0%	80.0%	100.0%
Sp	46.4%	28.0%	64.9%
PPV	50.0%	32.1%	67.9%
NPV	100.0%	76.9%	100.0% | 2) Serum tumor markers and ultrasound, T+ if serum markers above threshold or US “suspicious” postmenopausal women only |	Dis+	Dis-	Tot
---	---	---	---
T+	15	11	26
T-	1	18	19
Tot	16	29	45 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	93.8%	81.9%	100.0%
Sp	62.1%	44.4%	79.7%
PPV	57.7%	38.7%	76.7%
NPV	94.7%	84.7%	100.0%	 3) Fitted LR model using tumor markers and ultrasound, cut off probability of malignancy = 0.50; postmenopausal women only Same 2x2 table as #2 above **Results were reported, but have not been abstracted, for the following combinations:** PE + US (Premenopausal and all)	**Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: ? (references given for individual tests, not combinations) Sample size: +/- (only 45 menopausal) Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
					PE, tumor markers + US PE + tumor markers Tumor markers + US Postmenopausal women	
Schutter, Kenemans, Sohn, et al., 1994 **#940**	**Geographical location:** Amsterdam, The Netherlands **Dates:** Nov 1990-Dec 1992 **Size of population:** 276→ excluded 48 who did not meet inclusion criteria→ 228 **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, PPV, NPV, chi-square, Fishers exact, ROC curves, logistic regression **Blinding:** NR **Definition of positive and negative on screening test:** Logistic regression model: Exam: benign (0) or malignant (1)	**Age:** NR **Menopausal status (n [%]):** Post: 228 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Pelvic mass; age > 45; amenorrhea > 12 months; scheduled for surgery with biopsy or excision of mass; no history of BSO; no history of malignancy **Exclusion criteria:** Age ≥ 45; amenorrhea < 12 months; additional malignancy; physical exam "indeterminate;" no pre-operative CA-125	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 199/228 (87%) **Detected by imaging (n [%]):** 28/228 (12%) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Pelvic examination: abnormal if a mass distinguishable from the uterus was identified. Clinician was asked to characterize mass as benign (0) or malignant (1) 2) TVUS; used Finkler scoring system (ref#3, Table 3). A score of 7 or more was considered positive for malignancy (1). 3) Serum CA-125 >35 (1) is "malignant"	1) Physical exam, TVUS, and CA-125 all positive defines malignancy Dis+ / Dis- / Tot T+: 19 / 3 / 22 T-: 11 / 36 / 47 Tot: 30 / 39 / 69 Value / Lower 95% CI / Upper 95% CI Se: 62.0% / 44.6% / 79.4% Sp: 92.0% / 83.5% / 100.0% PPV: 86.4% / 72.0% / 100.0% NPV: 76.6% / 64.5% / 88.7% 2) Physical exam and TVUS both positive defines malignancy Dis+ / Dis- / Tot T+: 38 / 13 / 51 T-: 8 / 47 / 55 Tot: 46 / 60 / 106 Value / Lower 95% CI / Upper 95% CI Se: 83.0% / 72.1% / 93.9% Sp: 79.0% / 68.7% / 89.3% PPV: 74.5% / 62.5% / 86.5% NPV: 85.5% / 76.1% / 94.8% 3) TVUS and CA-125 positive defines malignancy Dis+ / Dis- / Tot T+: 20 / 5 / 25 T-: 12 / 38 / 50 Tot: 32 / 43 / 75 Lower / Upper	**Comments:** --Borderline tumors were excluded from 2x2 tables (included neither as benign nor malignant). --No specific cutoff is given or suggested using the new logistic regression model. --Reported values for N, Se, Sp, PPV, and NPV do not correspond precisely to 2x2 tables presented, presumably due to rounding error. **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: ? (unclear for physical exam; the other parameters used accepted/defined criteria) Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
	CA-125 ≤ 35 U/ml (0) or > 35 (1) Ultrasound score < 7 (0) or ≥ 7 (1)				Value 95% CI 95% CI Se **64.0%** 47.4% 80.6% Sp **89.0%** 79.6% 98.4% PPV 80.0% 64.3% 95.7% NPV 76.0% 64.2% 87.8% 4) Logistic regression model Z = -9.2378 + 2.2506(PE) + 1.6025(US) + 1.7293(CA-125). Probability of malignancy is 1/1+e exp(-z). "The ability to predict malignancy or benignancy of the pelvic mass appeared to be 81.5%". No AUC given. **Results were reported, but have not been abstracted, for the following combinations:** All possible combinations of ultrasound, CA-125, and exam findings.	

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Schutter, Sohn, Kristen, et al., 1998 **#730**	**Geographical location:** The Netherlands **Dates:** Nov 1990-Dec 1992 **Size of population:** 155 (151 could be classified as malignant or benign; 4 borderlines not included in analysis) **Registry** **Reference standard:** Histopathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, PPV, NPV, ROC curves, logistic regression **Blinding:** NR **Definition of positive and negative on screening test:** See Scoring column for individual tests. See Results column for logistic regression equation	**Age:** NR **Menopausal status (n [%]):** Post: 155 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients presenting with a pelvic mass who underwent surgery **Exclusion criteria:** Storage problems making sera unusable	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Physical exam: malignant or benign 2) Ultrasound: Finkler score* 7-10 positive, 1-6 negative 3) CA-125: ≥ 35 positive 4) CA72-4: ≥ 3 positive *Finkler score (Finkler et al 1988, ref #25) 1-cyst/smooth borders 2-cyst/irregular borders 3-cyst/echoes and irregular borders 4-6-equivocal 7-9-multiseptated or irregular cystic mass 10-pelvic mass with ascites	1) PE/US/CA-125/CA-72-4: Dis+ / Dis- / Tot T+ 27 / 0 / 27 T- 32 / 92 / 124 Tot 59 / 92 / 151 Value / Lower 95% CI / Upper 95% CI Se 46.0% / 33.3% / 58.7% Sp 100.0% / 96.7% / 100.0% PPV 100.0% / 88.9% / 100.0% NPV 74.2% / 66.5% / 81.9% 2) PE/US: Dis+ / Dis- / Tot T+ 47 / 12 / 59 T- 12 / 80 / 92 Tot 59 / 92 / 151 Value / Lower 95% CI / Upper 95% CI Se 80.0% / 69.8% / 90.2% Sp 87.0% / 80.1% / 93.9% PPV 79.7% / 69.4% / 89.9% NPV 87.0% / 80.1% / 93.8% 3) Logistic regression model: $1/(1+e^{-Z})$ where Z = -5.6816+2.2677*(US)-2.4928*(PE)+1.6057*(CA-125)+1.5866*(CA72-4) Dis+ / Dis- / Tot T+ 48 / 9 / 57 T- 11 / 83 / 94 Tot 59 / 92 / 151 Value / Lower 95% CI / Upper 95% CI Se 81.4% / 71.4% / 91.3% Sp 90.2% / 84.1% / 96.3% PPV 84.2% / 74.7% / 93.7%	**Comments:** --Borderline tumors not included (this is good) **Quality assessment:** Reference standard: + Verification bias: - (unclear how many negatives didn't have surgery) Test reliability/variability: + (US scoring system) Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

NPV 88.3% 81.8% 94.8%

4) US/CA-125/CA-72-4:

	Dis+	Dis-	Tot
T+	28	1	29
T-	31	91	122
Tot	59	92	151

	Value	Lower 95% CI	Upper 95% CI
Se	47.0%	34.3%	59.7%
Sp	99.0%	97.0%	100.0%
PPV	96.6%	89.9%	100.0%
NPV	74.6%	66.9%	82.3%

5) PE/US/CA-125:

	Dis+	Dis-	Tot
T+	34	5	39
T-	25	87	112
Tot	59	92	151

	Value	Lower 95% CI	Upper 95% CI
Se	57.6%	45.0%	70.2%
Sp	94.6%	89.9%	99.2%
PPV	87.2%	76.7%	97.7%
NPV	77.7%	70.0%	85.4%

Results were reported, but have not been abstracted, for the following combinations:
PE/US/CA-72-4
CA-125/CA-72-4
PE/CA-125/CA-72-4
PE/CA-125
PE/CA-72-4
US/CA-125
US/CA-72-4

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Strigini, Gadducci, Del Bravo, et al., 1996 **#4000**	**Geographical location:** Pisa, Italy **Dates:** Jan 1993-Jun 1994 **Size of population:** 109 total → 34 postmenopausal are reported here **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** See Scoring column	**Age:** Median: 43 Range: 18-80 **Menopausal status (n [%]):** Post: 34 (31%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Surgery scheduled for an adnexal mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Transvaginal ultrasound (TVUS) Classified masses as probably benign or malignant. Malignancy was defined as "solid portions with irregular structure, thick septae or papillae, irregular margins" 2) Doppler: pulsatility index < 1 was considered abnormal (reference #5 Kurjak). If no color flow was detected in the mass the Doppler was considered negative (benign). 3) CA-125 > 65 U/ml reference #11)	1) TVUS or CA-125 /postmenopausal only |	Dis+	Dis-	Tot
---	---	---	---
T+	12	3	15
T-	1	18	19
Tot	13	21	34 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	92.0%	77.3%	100.0%
Sp	86.0%	71.2%	100.0%
PPV	80.0%	59.8%	100.0%
NPV	94.7%	84.7%	100.0% | 2) TVUS and CA-125/postmenopausal only |	Dis+	Dis-	Tot
---	---	---	---
T+	6	1	7
T-	7	20	27
Tot	13	21	34 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	46.0%	18.9%	73.1%
Sp	95.0%	85.7%	100.0%
PPV	85.7%	59.8%	100.0%
NPV	74.1%	57.5%	90.6% | 3) TVUS or PI/postmenopausal only |	Dis+	Dis-	Tot
---	---	---	---
T+	12	5	17
T-	1	16	17
Tot	13	21	34 | |	Value	Lower 95% CI	Upper 95% CI
---	---	---	---
Se	92.0%	77.3%	100.0%
Sp	76.0%	57.7%	94.3%
PPV	70.6%	48.9%	92.2%
NPV	94.1%	82.9%	100.0%		**Comments:** --Definition of suspicious TVUS is fairly vague, subjective – no previously published standard is used --CA-125 > 65 U/ml also is not the usual cutoff, but a reference is given **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - (TVUS criteria vague) Sample size: - (only 34 postmenopausal women) Statistical tests: + Blinding: - Definition of +/- on screening test: +/- (TVUS criteria fairly vague) Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

4) TVUS and PI/postmenopausal only

	Dis+	Dis-	Tot
T+	9	1	10
T-	4	20	24
Tot	13	21	34

	Value	Lower 95% CI	Upper 95% CI
Se	69.0%	43.9%	94.1%
Sp	95.0%	85.7%	100.0%
PPV	90.0%	71.4%	100.0%
NPV	83.3%	68.4%	98.2%

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Tailor, Jurkovic, Bourne, et al., 1999 **#2910**	**Geographical location:** London, UK **Dates:** NR **Size of population:** 67 **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** ROC curves, logistic regression model using the network output gives probability of malignancy **Blinding:** NR **Definition of positive and negative on screening test:** Logistic regression model using the neural network output gives the probability of malignancy	**Age:** Mean: 39.6 Premenopausal (n = 45): mean age 38, range 20-52 Postmenopausal (n = 22): mean age 61, range 48-76 **Menopausal status (n [%]):** Pre (< 45): 45 (67%) Post (> 55): 22 (33%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Pelvic mass and scheduled for surgery **Exclusion criteria:** No histologic specimen obtained	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Age 2) Menopausal status 3) Tumor diameter 4) Tumor volume 5) Locularity 6) Presence of papillations 7) Echogenicity 8) Blood flow velocity waveforms 9) Peak systolic velocity 10) Time-average maximum velocity (TAMXV) 11) Pulsatility index 12) Resistance index Note: "best model" uses the following variables: age, maximum diameter, papillary projections, TAMXV	1) Neural network "best model" (entire data set); ROC AUC = 0.9987: Dis+ / Dis- / Tot T+: 14 / 0 / 14 T-: 1 / 52 / 53 Tot: 15 / 52 / 67 Value (Lower 95% CI, Upper 95% CI) Se: 93.3% (80.6%, 100.0%) Sp: 100.0% (94.2%, 100.0%) PPV: 100.0% (78.6%, 100.0%) NPV: 98.1% (94.5%, 100.0%) Training set n = 52 Test set n = 15 Data shown are for all cases, including training set	**Comments:** --Borderlines classified as malignant (not strictly true) --Very small sample size for modeling (n = 15 malignant cases in entire set) **Quality assessment:** Reference standard: - Verification bias: - (patients excluded who didn't have a pathology result [these could have been benign cases]) Test reliability/variability: - (don't know, other than small validation set) Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: + (authors used 15 patients for validation of the model found using the first 52)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Timmerman, Bourne, Tailor, et al., 1999 **#2930**	**Geographical location:** Belgium and UK **Dates:** Aug 1994-Jul 1996 **Size of population:** 191 **Prospective study of patients with a mass sent for U/S (consented prior to ultrasound and surgery)** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Multivariate logistic regression, ROC curves/AUC, Mann-Whitney U, student t, chi-square, Fisher exact **Blinding:** NR, but prospective **Definition of positive and negative on screening test:**	**Age:** Mean: 51.4 Benign (n = 140): mean (SD) = 49 (16) Malignant (n = 51): mean (SD) = 58 (14) **Menopausal status (n [%]):** Pre: 99 (51.8%) Post: 92 (48.2%) Benign (n = 140): 40% postmenopausal Malignant (n = 51): 71% postmenopausal **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** At least 1 adnexal mass; performance of TVUS with color Doppler; surgical removal of mass **Exclusion criteria:** Surgery not done	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**(A) RMI1**: calculated per Jacobs, Oram, Fairbanks, et al., 1990 (#6820) using US, menopausal status, and CA-125 **(B)Variables included in the final multivariable model**: 1) Papillary structures in mass > 3 mm 2) Serum CA-125 3) Color score (subjective from 1-4 depending on amount of blood flow) 4) Menopausal score (can't find exact criteria – unclear if same as criteria used for RMI1) **Multivariate model:** Probability of malignancy = 1/(1-e-z) Where z=(2.6369 x color score) + (0.0225 x CA-125) + (7.1062 x papillations > 3 mm score) + (2.6423 x postmenopausal score) - 13.6796 **(C) Morphologic Scoring system (Lerner, ref#18)**-details not given on how it was calculated	1) RMI1 with cutoff of 200; AUC = 0.882 Dis+ / Dis- / Tot T+: 33 / 11 / 44 T-: 16 / 113 / 129 Tot: 49 / 124 / 173 Value (Lower 95% CI, Upper 95% CI) Se: 67.3% (54.2%, 80.4%) Sp: 91.1% (86.1%, 96.1%) PPV: 75.0% (62.2%, 87.8%) NPV: 87.6% (81.9%, 93.3%) 2) Logistic regression model, n = 173 cutoff > 25% probability of malignancy; AUC = 0.967: Dis+ / Dis- / Tot T+: 47 / 16 / 63 T-: 2 / 108 / 110 Tot: 49 / 124 / 173 Value (Lower 95% CI, Upper 95% CI) Se: 95.9% (90.4%, 100.0%) Sp: 87.1% (81.2%, 93.0%) PPV: 74.6% (63.9%, 85.4%) NPV: 98.2% (95.7%, 100.0%) 3) Logistic regression model, n = 173, cutoff > 50% probability of malignancy, AUC = 0.967: Dis+ / Dis- / Tot T+: 43 / 10 / 53 T-: 6 / 114 / 120 Tot: 49 / 124 / 173 Value (Lower 95% CI, Upper 95% CI) Se: 87.8% (78.6%, 96.9%) Sp: 91.9% (87.1%, 96.7%)	**Comments:** --5 borderline tumors were counted as malignant **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: + Statistical tests: + Blinding: - (NR) Definition of +/- on screening test: + Explicit validation method?: +

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

PPV 81.1% 70.6% 91.7%
NPV 95.0% 91.1% 98.9%

4) Morphologic Index of Lerner (ref #18), AUC = 0.840:

	Dis+	Dis-	Tot
T+	49	55	104
T-	2	85	87
Tot	**51**	**140**	191

	Value	Lower 95% CI	Upper 95% CI
Se	**96.1%**	90.8%	100.0%
Sp	**60.7%**	52.6%	68.8%
PPV	47.1%	37.5%	56.7%
NPV	97.7%	94.6%	100.0%

Results were reported, but have not been abstracted, for the following combinations:
RMI1 for cutoffs of 50, 75, 100, 150 (n = 173 patients who had CA-125 available AUC = 0.882)
Timmerman model for training set (n = 116) and testing set (n = 57) but numbers NR for these subgroups

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Timmerman, Verrelst, Bourne, et al., 1999 **#5940**	**Geographical location:** Leuven, Belgium **Dates:** Aug 1994-Aug 1996 **Size of population:** 173 **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Logistic regression model, ROC curves/AUC **Blinding:** NR **Definition of positive and negative on screening test:**	**Age:** Mean: 53 Range: 22-93 **Menopausal status (n [%]):** Pre (< 45): 83 (48%) Post (> 55): 90 (52%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Adnexal mass; scheduled for surgery; had a CA-125 **Exclusion criteria:** No CA-125	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Neural Network 1 uses: menopause score (0 or 1), color score (1-4 subjective), papillations score (0 or 1), CA-125 2) Neural Network 2 uses: menopausal score (0 or 1), CA-125, ascites (0 or 1), unilocularity (0 or 1), papillations score (0 or 1), smooth walls score (0 or 1), bilaterality score (0 or 1), unilocular score (0 or 1) 3) Logistic Regression 2 uses: menopausal score (0 or 1), CA-125, ascites (0 or 1), unilocularity (0 or 1), smoothe walls (0 or 1), papillations (0 or 1), bilaterality (0 or 1) z=0.5948 meno + 0.0205CA-125 + 0.5446ascites - 0.762unilocular - 1.1606smoothe + 1.5049papillations + 0.7633bilateral - 1.0889 4) RMI1 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820]) uses: menopausal score, ultrasound score, CA-125	1) Neural Network 1 (AUC = 0.971) cutoff 45% probability of malignancy Dis+ / Dis- / Tot T+: 46 / 12 / 58 T-: 3 / 112 / 115 Tot: 49 / 124 / 173 Value (Lower 95% CI, Upper 95% CI) Se: 93.9% (87.2%, 100.0%) Sp: 90.3% (85.1%, 95.5%) PPV: 79.3% (68.9%, 89.7%) NPV: 97.4% (94.5%, 100.0%) 2) Neural Network 2 (AUC = 0.979) cutoff 60% probability of malignancy Dis+ / Dis- / Tot T+: 47 / 8 / 55 T-: 2 / 116 / 118 Tot: 49 / 124 / 173 Value (Lower 95% CI, Upper 95% CI) Se: 95.9% (90.3%, 100.0%) Sp: 93.5% (89.2%, 97.8%) PPV: 85.5% (76.1%, 94.8%) NPV: 98.3% (96.0%, 100.0%) 3) Logistic Regression 2 (AUC = 0.956) cutoff 60% probability of malignancy Dis+ / Dis- / Tot T+: 47 / 18 / 65 T-: 2 / 106 / 108 Tot: 49 / 124 / 173 Value (Lower 95% CI, Upper 95% CI) Se: 95.9% (90.3%, 100.0%) Sp: 85.5% (79.3%, 91.7%) PPV: 72.3% (61.4%, 83.2%)	**Comments:** --5 borderlines were counted as malignant **Quality assessment:** Reference standard: + Verification bias: - (only patients who had surgery are evaluated) Test reliability/variability: - Sample size: - Statistical tests: - Blinding: - (NR) Definition of +/- on screening test: + Explicit validation method?: +

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
					NPV 98.1% 95.6% 100.0%	
Tingulstad, Hagen, Skjeldestad, et al., 1996 **#3890**	**Geographical location:** Trondheim, Norway **Dates:** Feb 1992-Feb 1994 **Size of population:** 173 **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, chi-square, ROC curves **Blinding:** NR **Definition of positive and negative on screening test:** RMI1 = U x M x CA-125, where U = 0, 1, or 3; M = 1 or 3 RMI2 = U x M x CA-125, where U = 0, 1, or 4; M = 1 or 4	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 82 (47%) Post (> 55): 91 (53%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Admitted for surgery for a pelvic mass **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Ultrasound score (U) Based on presence of multilocular cystic lesions, solid areas, bilateral lesions, ascites, and intra-abdominal metastases scored 1 point each: If score = 0, then U = 0 If score = 1, then U = 1 If score ≥ 2, the U = 3 2) Menopause score (M): Premenopausal = 1 Postmenopausal = 3 3) CA-125 in kU/L RMI1 (Jacobs, Oram, Fairbanks, et al., 1990 [#6820]) RMI1 = M x U x CA-125 RMI2 (Tingulstad, Hagen, Skjeldestad, et al., 1996 [#3890]) RMI2 = M x U x CA-124 as RMI1 except U = 1 if total is 0 or 1 4 if total ≥ 2; M = 1 or 4	1) RMI1 cutoff 50: Dis+ Dis- Tot T+ 49 22 71 T- 7 95 102 Tot 56 117 173 Value Lower 95% CI Upper 95% CI Se 88.0% 79.5% 96.5% Sp 81.0% 73.9% 88.1% PPV 69.0% 58.3% 79.8% NPV 93.1% 88.2% 98.0% 2) RMI1 cutoff 100: Dis+ Dis- Tot T+ 44 9 53 T- 12 108 120 Tot 56 117 173 Value Lower 95% CI Upper 95% CI Se 79.0% 68.3% 89.7% Sp 92.0% 87.1% 96.9% PPV 83.0% 72.9% 93.1% NPV 90.0% 84.6% 95.4% 3) RMI1 cutoff 200: Dis+ Dis- Tot T+ 40 5 45 T- 16 112 128 Tot 56 117 173 Value Lower 95% CI Upper 95% CI Se 71.0% 59.1% 82.9% Sp 96.0% 92.4% 99.6% PPV 88.9% 79.7% 98.1% NPV 87.5% 81.8% 93.2%	**Comments:** --Borderline tumors counted as malignant **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: + Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: - (not for RMI2; this is a validation study of RMI1)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

4) RMI2 cutoff 100:

	Dis+	Dis-	Tot
T+	47	21	68
T-	9	96	105
Tot	**56**	**117**	173

	Value	Lower 95% CI	Upper 95% CI
Se	**84.0%**	74.4%	93.6%
Sp	**82.0%**	75.0%	89.0%
PPV	69.1%	58.1%	80.1%
NPV	91.4%	86.1%	96.8%

5) RMI2 cutoff 200:

	Dis+	Dis-	Tot
T+	45	9	54
T-	11	108	119
Tot	**56**	**117**	173

	Value	Lower 95% CI	Upper 95% CI
Se	**80.0%**	69.5%	90.5%
Sp	**92.0%**	87.1%	96.9%
PPV	83.3%	73.4%	93.3%
NPV	90.8%	85.6%	96.0%

Results were reported, but have not been abstracted, for the following combinations:
RMI1 and RMI 2 at the following cutoffs: 25, 50, 80, 100, 125, 150, 250

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Tingulstad, Hagen, Skjeldestad, et al., 1999 **#5950**	**Geographical location:** Trondheim, Norway **Dates:** Feb 1995-Jan 1997 **Size of population:** 365 **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** **Statistical tests used:** Chi-square, Mann-Whitney U, Se, Sp **Blinding:** NR **Definition of positive and negative on screening test:** RMI2 = M x U x CA-125, where U = 1 or 3; M = 1 or 3	**Age:** NR **Menopausal status (n [%]):** Pre (< 45): 193 (53%) Post (> 55): 172 (47%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age ≥ 30 with pelvic masses, scheduled for surgery **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Menopause score,(1 or 3) 2) Ultrasound score (1 or 3). Incorporates the following: multilocular, solid, bilateral, ascites, metastases 3) CA-125 DOUG: ABOVE LOOKS LIKE A DESCRIPTION OF RMI1????	1) RMI2 cutoff 100: Dis+ Dis- Tot T+ 58 61 119 T- 17 229 246 Tot 75 290 365 Value / Lower 95% CI / Upper 95% CI Se 77.0% 67.5% 86.5% Sp 79.0% 74.3% 83.7% PPV 48.7% 39.8% 57.7% NPV 93.1% 89.9% 96.3% 2) RMI2 cutoff 150: Dis+ Dis- Tot T+ 54 38 92 T- 21 252 273 Tot 75 290 365 Value / Lower 95% CI / Upper 95% CI Se 72.0% 61.8% 82.2% Sp 87.0% 83.1% 90.9% PPV 58.7% 48.6% 68.8% NPV 92.3% 89.1% 95.5% 3) RMI2 cutoff 200: Dis+ Dis- Tot T+ 53 23 76 T- 22 267 289 Tot 75 290 365 Value / Lower 95% CI / Upper 95% CI Se 71.0% 60.7% 81.3% Sp 92.0% 88.9% 95.1% PPV 69.7% 59.4% 80.1% NPV 92.4% 89.3% 95.4% 4) RMI2 cutoff 250:	**Comments:** --Borderline tumors counted as malignant **Quality assessment:** Reference standard: + Verification bias: - Test reliability/variability: + Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: + (this is a validation study)

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring

	Dis+	Dis-	Tot
T+	53	20	73
T-	23	270	293
Tot	75	290	366

	Value	Lower 95% CI	Upper 95% CI
Se	70.0%	59.6%	80.4%
Sp	93.0%	90.1%	95.9%
PPV	72.6%	62.4%	82.8%
NPV	92.2%	89.1%	95.2%

Results were reported, but have not been abstracted, for the following combinations:
RMI2 cutoff 50
RMI2 cutoff 300

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Torres, Derchain, Faundes, et al., 2002 **#2170**	**Geographical location:** São Paolo, Brazil **Dates:** Jan 1996-Mar 98 **Size of population:** 158 **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Logistic regression, ROC curves **Blinding:** NR **Definition of positive and negative on screening test:** RMI = U x M x CA-125 Results reported for a cutoff of 150	**Age:** NR **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Pelvic mass apparently restricted to adnexal; admitted for surgery **Exclusion criteria:** Known distant metastasis	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Ultrasound score (0-10) (Depriest et al 1993; Sassone et al., 1991) 0- unilocular simple cysts w/regular fine wall or lesion suggesting dermoid cyst 1-multilocular cyst w/regular and smooth wall (< 3 mm) or thick (> 3 mm) or solid homogeneous tumor w/hyperechogenic and well-defined wall 2-Unilocular cyst or multilocular cyst w/fine wall, with irregularity in the wall or septa (> 3 mm) 4-Multilocular cyst w/thick and irregular wall (irregularity in the wall or septa(> 3 mm), and/or irregular septa or cyst /papillary irregularity over 3 mm 5- Complex lesion w/irregularity in surface (< 3 mm) or badly-defined and irregular wall; or solid heterogeneous lesion 10-Complex lesion w/irregularity in surface (< 3 mm) or badly defined and irregular wall; or solid heterogenous lesion +1 ascites +2 wall expansive involvement > 3 mm 2) Menopausal status (1 or 3)	1) RMI (cutoff 150): Dis+ / Dis- / Tot T+ 53 / 19 / 72 T- 14 / 72 / 86 Tot 67 / 91 / 158 Value / Lower 95% CI / Upper 95% CI Se 79.0% / 69.2% / 88.8% Sp 79.0% / 70.6% / 87.4% PPV 73.6% / 63.4% / 83.8% NPV 83.7% / 75.9% / 91.5% AUC for RMI=0.90 2) RMI (cutoff 30): Dis+ / Dis- / Tot T+ 64 / 40 / 104 T- 3 / 51 / 54 Tot 67 / 91 / 158 Value / Lower 95% CI / Upper 95% CI Se 96.0% / 91.3% / 100.0% Sp 56.0% / 45.8% / 66.2% PPV 61.5% / 52.2% / 70.9% NPV 94.4% / 88.3% / 100.0% 3) RMI (cutoff 100): Dis+ / Dis- / Tot T+ 56 / 21 / 77 T- 11 / 70 / 81 Tot 67 / 91 / 158 Value / Lower 95% CI / Upper 95% CI Se 84.0% / 75.2% / 92.8% Sp 77.0% / 68.4% / 85.6% PPV 72.7% / 62.8% / 82.7% NPV 86.4% / 79.0% / 93.9% 4) RMI (cutoff 200):	**Comments:** --18 borderline tumors were treated as malignant (dubious) --The scoring system for ultrasound does not appear to be the same as that used for the other RMI studies **Quality assessment:** Reference standard: + Verification bias: - (don't know about patients who had masses/cysts, but didn't qualify for study) Test reliability/variability: - (not sure authors did the ultrasound scoring the way other published reports did) Sample size: + Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: - (unless this is a validation study [not sure they used the same method as DePriest et al (1993); Sassone et al., (1991) for ultrasound score])

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
				3) CA-125		

	Dis+	Dis-	Tot
T+	49	13	62
T-	18	78	96
Tot	67	91	158

	Value	Lower 95% CI	Upper 95% CI
Se	73.0%	62.4%	83.6%
Sp	86.0%	78.9%	93.1%
PPV	79.0%	68.9%	89.2%
NPV	81.3%	73.4%	89.1%

Results were reported, but have not been abstracted, for the following combinations:
RMI cutoff 500

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
Twickler, Forte, Santos-Ramos, et al., 1999 **#3080**	**Geographical location:** Dallas, TX **Dates:** Feb 1993-Aug 1996 **Size of population:** 304 with masses on exam: 217 had surgery 27 had sonographic followup 60 had no followup **Prospective study of patients with pelvic masses detected by exam** **Reference standard:** Pathology or clinical followup **Reference standard applied to all test negatives?:** No – 60 had no followup **Statistical tests used:** Mann-Whitney U, Fisher exact, linear logistic regression **Blinding:** NR **Definition of positive and negative on screening test:** Ovarian tumor index (see Scoring column for definition).	**Age:** Mean (SD): 38.6 (12.3) Range: 15-80 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Patients with clinically suspected adnexal masses **Exclusion criteria:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 304 (100%) **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Age (years) 2) Ovarian volume (ml) 3) Morphology scale (1-15) Sassone ref#3 4) PI (-10 x PI) 5) Vessel location: Peripheral -10 Central +10 6) Intense echoes -10 Ovarian Tumor Index is the sum of the above scores	Ovarian Tumor Index AUC=0.91 1) Ovarian Tumor Index cutoff 45: Dis+ / Dis- / Tot T+: 26 / 36 / 62 T-: 5 / 178 / 183 Tot: 30 / 214 / 245 Value (95% CI): Se 85.0% (72.2%–97.8%); Sp 83.0% (78.0%–88.0%); PPV 41.9% (29.7%–54.2%); NPV 97.3% (94.9%–99.6%) (likely rounding error in cell C) 2) Ovarian Tumor Index cutoff 55: Dis+ / Dis- / Tot T+: 23 / 19 / 42 T-: 7 / 195 / 202 Tot: 30 / 214 / 244 Value (95% CI): Se 78.0% (63.2%–92.8%); Sp 91.0% (87.2%–94.8%); PPV 54.8% (39.7%–69.8%); NPV 96.5% (94.0%–99.1%) 3) Ovarian Tumor Index cutoff 40: Dis+ / Dis- / Tot T+: 28 / 54 / 82 T-: 2 / 161 / 163 Tot: 30 / 214 / 245 Value (95% CI): Se 93.0% (83.9%–100.0%); Sp 75.0% (69.2%–80.8%); PPV 34.1% (23.9%–44.4%); NPV 98.8% (97.1%–100.0%)	**Comments:** --16 of 30 “malignant” neoplasms are borderline tumors (dubious) --No validation set was tested **Quality assessment:** Reference standard: - (no followup on 60 patients) Verification bias: - (no followup on 60 patients) Test reliability/variability: - Sample size: + Statistical tests: + Blinding: - (NR) Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
	Different cutoffs tested.				(likely rounding error)	
Valentin, Hagen, Tingulstad, et al., 2001 **#2340**	**Geographical location:** Malmo, Sweden and Trondheim, Norway **Dates:** NR **Size of population:** 157 original, 21 excluded due to no surgery, no pathology, etc.→ 136 included **Registry** **Reference standard:** Pathology **Reference standard applied to all test negatives?:** Yes **Statistical tests used:** Se, Sp, ROC curves **Blinding:** NR **Definition of positive and negative on screening test:** Tailor cutoff 50% Timmerman cutoff 50%	**Age:** "Slightly older" than population of Tailor et al. (1997) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Adnexal mass diagnosed clinically and scheduled for surgery **Exclusion criteria:** No surgery; no pathology from surgery available	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** 157 (100%) **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**A) Tailor method, cutoff 50% (Tailor et al 1997):** 1) Age 2) Menopausal status 3) Tumor diameter 4) Tumor volume 5) Locularity 6) Presence of papillations 7) Echogenicity 8) Blood flow velocity waveforms 9) Peak systolic velocity 10) Time-average maximum velocity 11) Pulsatility index 12) Resistance index **B) Timmerman method (ref #5):** 1) Papillary structures in mass > 3 mm 2) Serum CA-125 3) Color score (subjective from 1-4 depending on amount of	1) Tailor model, cutoff 50%; AUC=0.87: Dis+ / Dis- / Tot T+: 24 / 12 / 36 T-: 11 / 86 / 97 Tot: 35 / 98 / 133 Value / Lower 95% CI / Upper 95% CI Se: 69.0% / 53.7% / 84.3% Sp: 88.0% / 81.6% / 94.4% PPV: 66.7% / 51.3% / 82.1% NPV: 88.7% / 82.3% / 95.0% 2) Timmerman model, cutoff 50%; AUC = 0.84: Dis+ / Dis- / Tot T+: 18 / 11 / 29 T-: 11 / 42 / 53 Tot: 29 / 53 / 82 Value / Lower 95% CI / Upper 95% CI Se: 62.0% / 44.3% / 79.7% Sp: 79.0% / 68.0% / 90.0% PPV: 62.1% / 44.4% / 79.7% NPV: 79.2% / 68.3% / 90.2% **Results were reported, but have not been abstracted, for the following combinations:** Tailor model cutoff 25% Timmerman model cutoff 25%	**Comments:** --This is a validation study of two previously published models – both performed worse than in the original reports. Patient populations (numbers of different types of tumors) were different than in the original studies, possibly accounting for the worse performance. --Borderline tumors classified as malignant **Quality assessment:** Reference standard: + Verification bias: - (some were excluded because they did not have surgery) Test reliability/variability: + Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: + Explicit validation method?: +

Evidence Table 4 (continued)

Study	Study Design	Patients	Clinical Presentation	Items Included in Scoring System	Results	Comments/Quality Scoring
				blood flow) 4) Menopausal score (exact criteria unclear)		

Evidence Table 5: *Question 5: Among women with suspected benign masses on initial investigation, what are the sensitivity and specificity of monitoring with periodic CA-125 and/or interval ultrasound examinations for detecting malignant masses? How does the interval of testing/definition of change affect sensitivity and predictive value?*

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
Castillo, Alcazar, and Jurado, 2004 **#8040**	**Geographical location:** Pamplona, Spain **Dates:** Jan 1995-Jun 2002 **Size of population:** 8794 total; 215 had simple unilocular cysts **Screening study** **Reference standard:** Surgery or followup at 3 and 9 months **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Same examiner to eliminate interobserver variability **Statistical tests used:** Chi-square **Blinding:** No **Definition of positive and negative on screening test:** Increase in cyst size by 1 cm or more Simple cysts (sonolucent with wall	**Age:** Mean (SD): 59 (8.7) **Menopausal status (n [%]):** Post (> 55): 223 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Postmenopausal (at least 1 year without menses and older than 45, or hysterectomy with symptoms and older than 50) **Exclusion criteria:** Previous history of malignancy; history of bilateral oophorectomy **Loss to followup:** 66 (30.6%) (no statistically significant differences in any parameters between those who completed study and those who were lost to followup) Of remaining 149, 34 underwent immediate surgery.	**Symptomatic (n [%]):** 0 **Detected by exam (n [%]):** 26 (12%) **Detected by imaging (n [%]):** 215 (100%) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Monitoring test:** Ultrasound, CA-125 **Interval of testing:** 3 months, then every 6 months **Definition of change:** Progression: cyst increased in size at least 1 cm Regression: decrease at least 1 cm Resolution: cyst not detected in 2 consecutive exams	1) Cancer in patients followed: Dis+ / Dis- / Tot T+: 0 / 44 / 44 T-: 1 / 104 / 105 Tot: 1 / 148 / 149 Value / Lower 95% CI / Upper 95% CI Se: 0.0% / 0.0% / 0.0% Sp: 70.3% / 62.9% / 77.6% PPV: 0.0% / 0.0% / 0.0% NPV: 99.0% / 97.2% / 100.0%	**Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + Explicit validation method?: NA

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
	< 3 cm, no septations, solid areas, or papillary projections) less than 10 cm with CA-125 < 35, or patient choice, followed **Length of followup:** Median 27 months **Type of followup:** Repeat US and CA-125 **Followup interval:** 3 months, then 6 months					
Goldstein, Subra-manyam, Snyder, et al., 1989 **#10490**	**Geographical location:** New York, NY **Dates:** NR, but 3-year time period **Size of population:** 48; 16 followed **Case series** **Reference standard:** Surgery (26) or followup (16) **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No	**Age:** Range: 46-86 **Menopausal status (n [%]):** Post (> 55): 48 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Postmenopausal at least 12 months **Exclusion criteria:** NR **Loss to followup:** 6 (12.5%)	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Monitoring test:** Ultrasound **Interval of testing:** 3-6 months **Definition of change:** NR	1) Cancer in patients followed: Dis+ / Dis- / Tot T+: 0 / 2 / 2 T-: 0 / 14 / 14 Tot: 0 / 16 / 16 Value / Lower 95% CI / Upper 95% CI Se: - / - / - Sp: 87.5% / 71.3% / 100.0% PPV: 0.0% / 0.0% / 0.0% NPV: 100.0% / 78.6% / 100.0%	**Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: - Explicit validation method?: NA

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
	Statistical tests used: None **Blinding:** "Treatment rationale was not always available to the sonographers." **Definition of positive and negative on screening test:** Unilateral, no septations or solid components, diameter ≤ 5 cm, no ascites **Length of followup:** Mean 29 months (range 10-73 months) **Type of followup:** Repeat transabdominal US **Followup interval:** 3-6 months					

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
Kurjak, Shalan, Kupesic, et al., 1994 **#4470**	**Geographical location:** Zagreb, Croatia **Dates:** Jan 1988-Dec 1992 **Size of population:** 5013 in initial screening population 404 had simple cysts; results only reported for 88 with 2nd scan **Screening study** **Reference standard:** Surgery or followup **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not referenced **Statistical tests used:** None **Blinding:** No **Definition of positive and negative on screening test:** Cyst diameter ≥ 2.5 cm, < 5 cm , with resistance index ≥ 0.41 **Length of followup:** 6 months **Type of followup:**	**Age:** NR for group with cyst **Menopausal status (n [%]):** Pre (< 45): 280 (69.3%) Peri (45-55): 0 Post (> 55): 124 (30.6%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** 40 years or older, absence of symptoms **Exclusion criteria:** NR **Loss to followup:** NR; apparently 0	**Symptomatic (n [%]):** 0 **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Monitoring test:** Repeat ultrasound **Interval of testing:** 6 months **Definition of change:** Not explicitly defined	1) Cancer during followup after initial benign: Dis+ / Dis- / Tot T+: 1 / 17 / 18 T-: 0 / 69 / 69 Tot: 1 / 86 / 87 Value (Lower 95% CI, Upper 95% CI) Se: 100.0% (0.0%, 100.0%) Sp: 80.2% (71.8%, 88.6%) PPV: 5.6% (0.0%, 16.1%) NPV: 100.0% (95.7%, 100.0%)	**Comments:** --Incomplete details on other 316 subjects --Short duration of followup **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: + Explicit validation method?: NA

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
	Repeat ultrasound **Followup interval:** 6 months					
Levine, Gosink, Wolf, et al., 1992 **#10320**	**Geographical location:** Portland, OR **Dates:** Oct 1989-June 1990 **Size of population:** 184; 32 had simple cysts at initial evaluation, 31 developed over course of study **Other:** Cross-sectional, volunteer screening **Reference standard:** Followup **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Not discussed or referenced **Statistical tests used:** Chi-square, t-test **Blinding:** No **Definition of positive and negative on screening test:** Simple cyst –	**Age:** Mean (SD): 63.6 (8.1) Range: 50-85 **Menopausal status (n [%]):** Post (> 55): 184 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age ≥ 59, postmenopausal at least 1 year **Exclusion criteria:** NR **Loss to followup:** 49 of 63 women with cysts had ultrasound followup (22.2% loss to followup)	**Symptomatic (n [%]):** 0 **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Monitoring test:** Transabdominal/ transvaginal ultrasound Doppler and CA-125 if abnormal **Interval of testing:** 3 months **Definition of change:** Increased: diameter change ≥ 3 mm Decreased: diameter change ≤ 3 mm	1) Cancer in simple cysts: Dis+ / Dis- / Tot T+: 0 / 0 / 0 T-: 0 / 32 / 32 Tot: 0 / 32 / 32 Value / Lower 95% CI / Upper 95% CI Se: - / - / - Sp: 100.0% / 90.6% / 100.0% PPV: - / - / - NPV: 100.0% / 90.6% / 100.0%	**Comments:** --Results not presented clearly --Inconsistent followup strategy **Quality assessment:** Reference standard: - Verification bias: - Test reliability/variability: - Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: + Explicit validation method?: NA

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
	completely anechoic, unilocular, nonseptated **Length of followup:** 2 years ("over half one year or more") **Type of followup:** Transabdominal and transvaginal ultrasound **Followup interval:** 3 months x 1 year, then every 6 months					
Maggino, Gadducci, D'Addario, et al., 1994 **#4500**	**Geographical location:** Padua, Pisa, Bari, Brescia, and Milan, Italy **Dates:** Mar 1991-Mar 1992 **Size of population:** 335; 45 with benign cyst and CA-125 < 35 **Case series** **Reference standard:** Surgery or followup **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** Se, Sp, PPV, NPV, Kappa	**Age:** NR for subgroup of interest **Menopausal status (n [%]):** Post (> 55): 45 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Pelvic mass, at least 1 year post-menopausal **Exclusion criteria:** Incomplete US data, no CA-125, no histology for patients with surgery **Loss to followup:**	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Monitoring test:** Not systematic **Interval of testing:** NR **Definition of change:** NR	1) Cancer during followup after initial benign: Dis+ / Dis- / Tot T+: 0 / 0 / 0 T-: 0 / 45 / 45 Tot: 0 / 45 / 45 Value / Lower 95% CI / Upper 95% CI Se: - / - / - Sp: 100.0% / 93.3% / 100.0% PPV: - / - / - NPV: 100.0% / 93.3% / 100.0%	**Comments:** --No details on followup strategy --No details on length of followup **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: + Sample size: - Statistical tests: - Blinding: - Definition of +/- on screening test: + Explicit validation method?: NA

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
		2/45 (4.4%)				
	Blinding: No **Definition of positive and negative on screening test:** < 5 cm, thin wall, no echoes, ≤ 3 thin septa, no free fluid in pelvis **Length of followup:** NR **Type of followup:** NR **Followup interval:** NR					
Menon, Talaat, Rosenthal, et al., 2000 **#2780**	**Geographical location:** United Kingdom **Dates:** 1986-1989 **Size of population:** 22,000 screened; 741 with elevated CA-125 had US; 97 scans were abnormal or equivocal **Screening study** **Reference standard:** Pathology and clinical followup **Reference standard applied to all test negatives?:** No **Test reliability established?:** ?	**Age:** NR **Menopausal status (n [%]):** All older than 45 (22,000/100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Age > 45, CA-125 > 30 on screening **Exclusion criteria:** Premenopausal, CA-125 < 30 **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Monitoring test:** Pelvic ultrasound/serial CA-125 **Interval of testing:** Depends on findings. If equivocal US, repeat every 6 weeks until normal or abnormal. If normal, repeat the CA-125 every 3 months for a year. If US abnormal, refer to gynecologist for a decision. **Definition of nl or abnl:** Normal: volume < 8.8 ml or ovaries not visualized Equivocal: volume < 8.8 with abnormal morphology Abnormal: volume >	Among 17 patients with an equivocal scan who were triaged to followup in 4-6 weeks: -9 had simple cysts, did not have surgery, and did not develop cancer (true negatives); -1 died of ovarian cancer before her repeat ultrasound (can't categorize); -1 died of pneumonia before a repeat ultrasound could be done (can't categorize); -5 had surgery with benign pathology found (number of followup scans not specified; all false positives); -1 had surgery with ovarian cancer found (true positive).	**Comments:** --Question 4 refers to patients with masses thought to be benign. The only categories in this paper are normal (doesn't differentiate those with probably benign masses), equivocal (can't classify as normal or abnormal) and abnormal. Not sure the "equivocal" category can be considered "masses thought to be benign," but that is the group we focused on to answer Question 4. --Median followup was excellent but the exact number lost to followup was not given. --Authors didn't specify what happens if a patient with a normal scan subsequently develops a rising CA-125. **Quality assessment:** Reference standard: +/- Verification bias: +/- (median followup was specified but not

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
	Statistical tests used: **Blinding:** No **Definition of positive and negative on screening test:** Yes **Length of followup:** Median 6.8 years **Type of followup:** Annual questionnaires, Tumor Registry, pathology **Followup interval:** If scan normal, repeat CA-125 every 3 months for a year. If scan equivocal, repeat every 6 weeks until it can be classified as normal or abnormal. If scan abnormal, refer to gynecologist – surgical intervention at the gynecologist's discretion.			8.8 irrespective of morphology		all negatives got surgery) Test reliability/variability: Sample size: + Statistical tests: ? Blinding: - Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
Modesitt, Pavlik, Ueland, et al., 2003 **#5560**	**Geographical location:** Lexington, KY **Dates:** 1987-2002 **Size of population:** 15,106 screened with TVUS; 2763 women had 3259 unilocular ovarian cysts **Screening study** **Reference standard:** Clinical followup (mean 6.3 years) **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** Yes **Statistical tests used:** **Blinding:** **Definition of positive and negative on screening test:** Negative: unilocular simple cyst Positive: any septum, solid area, papillary projections, or volume > 10 cm^3 **Length of followup:** 6.3 years **Type of followup:**	**Age:** NR **Menopausal status (n [%]):** Post (> 55): 100% **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** 49% of patients with unilocular cysts were on hormone replacement therapy **Inclusion criteria:** At least 50 years old **Exclusion criteria:** Known ovarian tumor prior to screening, previous diagnosis of ovarian cancer, symptoms consistent with a pelvic mass **Loss to followup:** NR; mean followup 6.3 years	**Symptomatic (n [%]):** 0 **Detected by exam (n [%]):** 0 **Detected by imaging (n [%]):** 2763 (100%) **Combination (n [%]):** 0 **Additional data used for diagnosis:** NR	**Monitoring test:** TVUS **Interval of testing:** Repeat in 4-6 weeks with Doppler and CA-125 if abnormality detected. A tumor score from 0-10 was assigned based on volume and structure at the second scan. If cyst still appears simple at second scan, repeat TVUS every 3-6 months. **Definition of change:** Development of a septum or solid area, volume >10 cm^3, papillary projections into cyst	2763 women had 3259 unilocular cysts. Of these, 2261 (69.4%) spontaneously resolved, 220 (6.8%) persisted, 726 (22.3%) developed changes such as septa or solid areas and were no longer considered simple, 12 (0.3%) had an ovary that couldn't be visualized on a subsequent scan, and 40 (1.2%) were removed during a subsequent surgery. Authors did not report how many patients had surgery. They did report that **10 patients in this "unilocular cyst" population subsequently developed ovarian cancer**: -7 of these had an additional abnormal area that developed on their TVUS besides the simple cyst (so presumably they are not "false negatives"-they were caught by screening eventually). -2 had the cyst in question resolve spontaneously, but were ultimately diagnosed with ovarian cancer (should these be considered false negatives – probably not). -1 was ultimately diagnosed with cancer in the opposite ovary (probably not a false negative). They therefore claim **a 0% false negative rate in unilocular cysts < 10 cm (none of these cysts subsequently turned into cancer**). If all 3 in question above are treated as false negatives, the 2x2 table is below: ("by patient," not "by cyst") Dis+ Dis- Tot T+ 7 0 7 T- 3 2753 2756 Tot 10 2753 2763	**Comments:** --Good evidence that simple ovarian cysts almost never represent cancer and can be followed. --Could argue that 3 of the patients who developed cancer should be treated as "false negatives" but the cancers developed either after the original cyst had resolved or in the other ovary – did 2x2 tables for both interpretations. **Quality assessment:** Reference standard: + Verification bias: + (good followup) Test reliability/variability: + Sample size: + Statistical tests: - (none) Blinding: Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring

Followup interval:

	Value	Lower 95% CI	Upper 95% CI
Se	70.0%	41.6%	98.4%
Sp	100.0%	99.9%	100.0%
PPV	100.0%	57.1%	100.0%
NPV	99.9%	99.8%	100.0%

2) On a “by cyst” basis, there are no false negatives, since the 3 patients who had cancers did so after the cysts in question resolved (or in the opposite ovary):

	Dis+	Dis-	Tot
T+	7	0	7
T-	0	3252	3252
Tot	7	3252	3259

	Value	Lower 95% CI	Upper 95% CI
Se	100.0%	57.1%	100.0%
Sp	100.0%	99.9%	100.0%
PPV	100.0%	57.1%	100.0%
NPV	100.0%	99.9%	100.0%

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
Schincaglia, Brondelli, Cicognani, et al., 1994 **#4520**	**Geographical location:** Bologna, Italy **Dates:** Aug 1988-Jun 1992 **Size of population:** -3541 screened -347 were asked to followup, and can be split into 2 groups: -249 were followed with additional scans but not deemed "abnormal" enough to refer for FNA/surgery -98 were referred for possible FNA ("abnormal") -2 cancers diagnosed in the 98 who were referred -no cancers among the rest of the population (maximum followup was 1 year) **Screening study** **Reference standard:** Pathology and clinical followup up to 1 year **Reference standard applied to all test negatives?:** Yes **Test reliability established?:** No **Statistical tests used:** SP, PPV	**Age:** NR **Menopausal status (n [%]):** Post (> 55): 3541 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Postmenopausal, no prior pelvic surgery or pelvic symptoms **Exclusion criteria:** Prior pelvic surgery, pelvic symptoms **Loss to followup:** Not stated, but followup listed as "at least one year" for negative screens and those managed conservatively	**Symptomatic (n [%]):** 0 **Detected by exam (n [%]):** 0 **Detected by imaging (n [%]):** All who qualified for followup (347) **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Monitoring test:** Ultrasound **Interval of testing:** See below **Definition of change:** See below Volume of ovaries: < 9 negative < 9 with a cyst: followup US in 6 months. Increased volume → refer to level II 9-15: followup 3 and 6 months. Unchanged → refer to level II > 15 referral to level II Level II: Morphology assessment and biopsy if feasible Surgery if FNA not feasible, inadequate, positive, or patient refuses FNA	1) Of the 347 patients selected for followup initially, 283 were deemed appropriate for followup using repeat ultrasound at 3- to 6-month intervals without immediate referral for FNA/surgery. Of these 283, 34 subsequently had concerning US results and were referred for level II scan and possible FNA. The results of this group of 34 are not given separately. Of the 249 who had non-concerning followup scans, none developed cancer, with followup of "at least" 1 year. Therefore Specificity is 100% for patients with an initial abnormal but "probably benign" finding who had reassuring followup studies. Sensitivity within this group cannot be calculated with the information given. Dis+ / Dis- / Tot T+: 2 / 96 / 98 T-: 0 / 249 / 249 Tot: 2 / 345 / 347 Value (Lower 95% CI, Upper 95% CI) Se: 100.0% (0.0%, 100.0%) Sp: 72.2% (67.4%, 76.9%) PPV: 2.0% (0.0%, 4.8%) NPV: 100.0% (98.8%, 100.0%)	**Comments:** --The category "probably benign" is not really defined here, so abstractor chose the patients who had something abnormal on their original screen. --According to the authors, none of the patients who were not referred to level 2 developed ovarian cancer, but the followup time was fairly minimal compared to other large screening studies. **Quality assessment:** Reference standard: + Verification bias: - (followup not very long) Test reliability/variability: + Sample size: + Statistical tests: - Blinding: - (NR) Definition of +/- on screening test: + Explicit validation method?: -

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
	Blinding: NR **Definition of positive and negative on screening test:** See under "Monitoring Strategy" **Length of followup:** At least 1 year **Type of followup:** Questionnaire for negative screens; others had followup studies as specified under "Monitoring Strategy" **Followup interval:** Varied by ultrasound findings					
Valentin and Akrawi, 2002 **#8490**	**Geographical location:** Malmo, Sweden **Dates:** June 1991-Nov 2000 **Size of population:** 162; 134 agreed to followup (28 not followed older, had higher mortality) **Consecutive case-series** **Reference standard:** Surgery or followup **Reference standard applied to all test**	**Age:** Median: 61 Range: 47-87 **Menopausal status (n [%]):** Post (> 55): 134 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Inclusion criteria:** Cysts "judged to be benign," age ≥ 40, menopausal for at least 1 year	**Symptomatic (n [%]):** 84 (62.7%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	**Monitoring test:** TVUS **Interval of testing:** 3, 6, 12, then every 12 months **Definition of change:** Positive difference between largest diameter at most recent examination and at initial examination, or cyst "more complex" Morphology improved if cyst complexity decreased	1) Change in US as positive test, cancer as disease: Dis+ / Dis- / Tot T+: 0 / 7 / 7 T-: 0 / 127 / 127 Tot: 0 / 134 / 134 Value / Lower 95% CI / Upper 95% CI Se Sp: 94.8% / 91.0% / 98.5% PPV: 0.0% / 0.0% / 0.0% NPV: 100.0% / 97.6% / 100.0% 4 additional patients were operated on for other causes	**Comments:** --Complete followup **Quality assessment:** Reference standard: + Verification bias: + Test reliability/variability: - Sample size: - Statistical tests: + Blinding: - Definition of +/- on screening test: + (not defined, but references provided) Explicit validation method?: NA

Evidence Table 5 (continued)

Study	Study Design	Patients	Clinical Presentation	Monitoring Strategy	Results	Comments/Quality Scoring
	negatives?: Yes **Test reliability established?:** Yes **Statistical tests used:** Sp, PPV, NPV **Blinding:** No **Definition of positive and negative on screening test:** Referenced but not explicitly described in this paper **Length of followup:** Median 3 years (range 4 months-8 years) **Type of followup:** Transvaginal ultrasound **Followup interval:** 3, 6, 12 months, then every 12 months	**Exclusion criteria:** NR **Loss to followup:** 0% (mortality and surgical data obtained from Swedish national registries)				

Evidence Table 6: *Question 6: Among women with adnexal masses, what are the morbidity and mortality from diagnostic surgery (laparoscopy or laparotomy)? At what point does the risk of surgery outweigh the risk of detecting malignancy?*

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Canis, Mage, Pouly, et al., 1994 **#4610**	**Geographical location:** Clermont-Ferrand, France **Dates:** NR **Size of population:** 757 **Registry** **Morbidity definitions:** Not described **Length of followup after surgery:** Mean 42 months (range 3-153)	**Age:** Mean (SD): 35.8 (12.6) Range: 8-84 **Menopausal status (n [%]):** Pre (< 45): 671 (88.6%) Post (> 55): 86 (11.4%) 92 patients > 50 **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** 81/620 (13.1%)	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Ultrasound Age CA-125 "Recurrent cysts" < endometriomas, paraovarian and functional cysts, pseudoperitoneal cysts, and hydrosalpinges excluded; not clear if excluded before or after surgical diagnosis	1) Mortality: 0/757 (95% CI, 0 to 0.6%) 2) Morbidity (total all complications): 8/727 (1.1%) 3) Specific complications: 1 gastric laceration 1 acute abdomen 1 peritonitis (sigmoid laceration) 2 ovarian abscess 2 peritonitis (ruptured teratoma) 1 led to immediate re-operation 1 led to operation for CPP 12 mo later 1 abdominal wall endometrioma re-operated 4) Rate of conversion to laparotomy: Not described for non-malignant tumors	**Comments:** --Unclear whether some benign cases excluded after surgery **Quality assessment:** Size of population from which sample drawn: - Number of cases: + Patient selection: - Application of reference standard: -

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Canis, Mashiach, Wattiez, et al., 2004 **#7720**	**Geographical location:** Clermont-Ferrard, France **Dates:** ?1992-1997 (specific dates not given) **Size of population:** 839 total , 141 with frozen section **Single center** **Morbidity definitions:** Sensitivity/specificity of frozen section **Length of followup after surgery:** Up to 10 years (not uniformly)	**Age:** Mean (SD): 43.6 (15.9) **Menopausal status (n [%]):** Pre (< 45): 99 (70.2%) Post (> 55): 42 (29.8%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: NR 2) Morbidity (total all complications): Sensitivity/specificity of frozen section: Low malignant potential = cancer, "unclear" results on frozen = positive test Dis+ / Dis- / Tot T+: 47 / 4 / 51 T-: 4 / 86 / 90 Tot: 51 / 90 / 141 Value (Lower 95% CI, Upper 95% CI) Se: 92.2% (84.8%, 99.5%) Sp: 95.6% (91.3%, 99.8%) PPV: 92.2% (84.8%, 99.5%) NPV: 95.6% (91.3%, 99.8%) 3) Specific complications: NR 4) Rate of conversion to laparotomy: NR	**Comments:** --Clinical history prior to surgery not described **Quality assessment:** Size of population from which sample drawn: + Number of cases: - Patient selection: Not described Application of reference standard: +
Carley, Klingele, Gebhart, et al., 2002 **#8500**	**Geographical location:** Rochester, MN **Dates:** Dec 1995-Nov 2000 **Size of population:** 106 **Single center** **Morbidity definitions:** -Febrile morbidity -Transfusion -Conversion **Length of followup after surgery:** Not specified	**Age:** Mean (SD): Laparotomy: 46.4 (13.5) Laparoscopy: 49.2 (15.9) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** History not reported **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Masses > 7 cm or with CA-125 > 35 not included in analysis; malignant pathology excluded	1) Mortality: NR 2) Morbidity (total all complications): Laparotomy: 2/44 (4.6%; 95% CI, 0.7, 16.7%) 3) Specific complications: Transfusion: Laparotomy: 1/44 (2.3%; 95% CI, 0 to 13.5%) Laparoscopy (including conversions): 0/62 (0 to 8.6%) Febrile morbidity: Laparotomy: 1/44 (2.3%; 0 to 13.5%) Laparoscopy: 0/62 (0 to 8.6%) 4) Rate of conversion to laparotomy: 16%	**Quality assessment:** Size of population from which sample drawn: + Number of cases: - Patient selection: - Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Chapron, Dubuisson, and Capella-Allouc, 1997 **#6150**	**Geographical location:** Paris, France **Dates:** Jan 1989-Dec 1994 **Size of population:** 186 **Single center** **Morbidity definitions:** -Conversion to laparotomy -Complications **Length of followup after surgery:** NR	**Age:** Mean: Laparotomy 45.7 Laparoscopy 49.5 Range: Laparotomy 18-72 Laparoscopy 19-82 **Menopausal status (n [%]):** Laparotomy Pre (< 40): 18 (27.7%) Peri (40-50): 28 (43.1%) Post (> 50): 19 (29.2%) Laparoscopy Pre (< 40): 21 (18.2%) Peri (40-50): 38 (31.4%) Post (> 50): 61 (50.4%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** All patients underwent removal of adnexal Patients with "prophylactic" adnexal removal due to age, family history excluded	1) Mortality: Laparotomy: 0/65 (95% CI, 0 to 8.3%) Laparoscopy: 0/121 (0 to 4.5%) 2) Morbidity (total all complications): Laparotomy: 10/65 (15.4%; 95% CI, 8.9 to 27.0%) Laparoscopy: 10/121 (8.3%; 4.6 to 15.0%) 3) Specific complications: Laparotomy: Cystitis: 2/65 Febrile morbidity: 2/65 Abdominal wall hematoma: 1/65 Abdominal wall abscess: 2/65 Bowel obstruction: 1/65 Evisceration: 2/65 Laparoscopy: Urinary tract infection: 1/121 Febrile morbidity: 5/121 Bowel obstruction: 1/121 Evisceration: 2/121 (both re-operated) Sigmoid injury: 1/121 (re-operated) 4) Rate of conversion to laparotomy: 19/140 (13.6%; 95% CI, 9.0 to 20.6%)	**Comments:** --Groups not comparable in terms of baseline assessment (those with higher suspicion of malignancy went to directly to laparotomy – and 21 laparotomy patients were emergency surgery secondary to "considerable hemoperitoneum") --Description of clinical pathway to surgery not described --19 of the "laparotomy" group started as laparoscopy; data not provided to summarize results by "intention to treat" --Reoperation for laparotomy group not mentioned --Didn't include transfusion as specific com[placation (1/65 in laparotomy) **Quality assessment:** Size of population from which sample drawn: - (referral base not described) Number of cases: - Patient selection: - (not described) Application of reference standard: + (all patients underwent surgery)

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Chi, Abu-Rustum, Sonoda, et al., 2004 **#7870**	**Geographical location:** New York, NY **Dates:** Jan 1991-Dec 2000 **Size of population:** 1451 (146 with diagnostic laparoscopy) **Single center** **Morbidity definitions:** -Grade 1: use of oral medications, bedside interventions -Grade 2: IV medications, TPN, enteral nutrition, or blood transfusion -Grade 3: interventional radiology, endoscopy, intubation, or operation -Grade 4: Residual and lasting disability that requires major rehabilitation or organ resection -Grade 5: Death **Length of followup after surgery:** 30 days	**Age:** (all patients) Median: 54 Range: (15-88) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** Pre-procedure history not reported **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	Results presented for diagnostic laparoscopy only 1) Mortality: 3/146 (2.5%; 95% CI, 0.5 to 6.3%) 2) Morbidity (total all complications): 19/146 (13.0%; 95% CI, 8.6 to 19.8%) 3) Specific complications: Grade 1: 14/146 (9.6%; 95% CI, 5.8 to 15.8%) Grade 2: 1/146 (0.7%; 0 to 4.3%) Grade 3: 4/146 (2.7%; 0.9 to 7.2%) Grade 4: 0/146 (0%; 0 to 3.76%) 4) Rate of conversion to laparotomy: 15/146 (10.3%; 95% CI, 6.4 to 16.6%) 5) Multivariate analysis: Complications significantly more likely with older age, history of radiation therapy, malignancy	**Comments:** --Gynecological oncology service --Proportion of patients with adnexal mass (as opposed to other malignancies) not reported **Quality assessment:** Size of population from which sample drawn: + Number of cases: - Patient selection: - Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Childers, Nasseri, and Surwit, 1996 **#6940**	**Geographical location:** Tucson, AZ **Dates:** July 1991-Jan 1995 **Size of population:** 138 **Single center** **Morbidity definitions:** Complications **Length of followup after surgery:** 23-50 months (mean 37) in patients with malignancy; not given for others	**Age:** Mean: 52 Range: 9-91 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** "Suspicious adnexal mass", which on ultrasound did not meet all criteria: --Size < 10 cm --Unilateral --Smooth borders --No excrescences --No solid parts --No fluid in cul-de-sac, and with no ascites, malignant cells on paracentesis, or upper abdominal masses	1) Mortality: 0/138 (95% CI, 0 to 4.0%) 2) Morbidity (total all complications): 14/138 (10.1%; 95% CI, 6.2 to 16.7%) 3) Specific complications: Enterotomy: 1/138 Vena Cava injury: 1/138 Bowel herniation: 1/138 (re-operated) Febrile morbidity: 2/138 Ileus: 3/138 Cardiac arrhythmia: 4/138 Urinary retention: 2/138 4) Rate of conversion to laparotomy: 11/138 (8.0%; 95% CI, 4.5 to 14.1%) All for either technical reasons, debulking, or staging; none for complications 5) Malignancy 19/138 (13.8%; 95% CI, 9.1 to 20.9%)	**Comments:** --Selected population – gynecological oncology service, higher probability of malignancy --No data on initial clinical presentation --Results not stratified by age or menopausal status **Quality assessment:** Size of population from which sample drawn: - (no data on referral base) Number of cases: - (< 200) Patient selection: - (unclear how many patients did not undergo laparoscopy) Application of reference standard: + (all underwent pathology)

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Deckardt, Saks, and Graeff, 1994 **#4310**	**Geographical location:** Munich, Germany **Dates:** NR **Size of population:** 192 **Single center randomized trial** **Morbidity definitions:** Complications **Length of followup after surgery:** NR	**Age:** Laparotomy: Mean: 43.6 Range: 20-84 Laparoscopy: Mean: 40.1 Range: 18-74 **Menopausal status (n [%]):** Laparotomy: Pre (< 45): 73.7% Post (> 55): 26.3% Laparoscopy: Pre (< 45): 77.6% Post (> 55): 22.4% **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Premenopausal: mass > 6 cm, persistent after 6 months, symptomatic; infertility; not a corpus luteum cyst by ultrasound Postmenopausal: any adnexal mass Patients randomized according to ward where they were admitted	1) Mortality: Laparotomy: 0/76 (95% CI, 0 to 7.1%) Laparoscopy: 0/116 (0 to 4.7%) 2) Morbidity (total all complications): Laparotomy: 23/76 (30.3%; 21.8 to 42.3%) Laparoscopy: 13/116 (11.2%; 6.8 to 18.7%) 3) Specific complications: Bladder injury: Laparotomy 1/76, laparoscopy 0/116 Incisional hernia: Laparotomy 3/76, laparoscopy 1/116 Umbilical hernia: Laparotomy 0/76, laparoscopy 1/116 UTI: Laparotomy 12/76, laparoscopy 2/116 Febrile morbidity: Laparotomy 3/76, laparoscopy 2/116 Bowel obstruction: Laparotomy 1/76, laparoscopy2/116 Chemical peritonitis: Laparotomy 0/76, laparoscopy 1/116 Small bowel injury: Laparotomy 2/76, laparoscopy 1/116 Pulmonary embolus: Laparotomy 1/76, laparoscopy 0/116 Wound dehiscence: Laparotomy 2/76, laparoscopy 1/116 4) Rate of conversion to laparotomy: 4/116 (3.5%; 1.2 to 9.0%) 5) 46% of laparotomy patients received prophylactic antibiotics, compared to 2.6% of laparoscopy patients 6) Laparoscopy patients significantly more likely to have cystectomy (60.0% vs 20.2%), less likely to have oophorectomy (0.8% vs 20.2%), less likely to have bilateral salphingo-oophorectomy (4.0 vs 21.4%) 7) Reoperation 1/76 for laparotomy (assume 0/116 for laparoscopy?)	**Comments:** --Randomization not well described; differences in baseline characteristics, types of procedure performed suggest some bias in treatment allocation --No significance testing between groups to evaluate randomization **Quality assessment:** Size of population from which sample drawn: - (not described) Number of cases: - Patient selection: - (possible bias in treatment allocation) Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Dottino, Levine, Ripley, et al., 1999 **#6920**	**Geographical location:** New York, NY **Dates:** Apr 1992-Apr 1996 **Size of population:** 160 **Single center** **Morbidity definitions:** -Conversion to laparotomy -"Complications" -Reoperation -Misdiagnosis **Length of followup after surgery:** NR, but paper published Feb 1999, last patient enrolled Apr 1996	**Age:** Mean (SD): 52.2 (13.1) **Menopausal status (n [%]):** Pre (< 45): 75 (47%) Post (> 55): 85 (53%) **Race/ethnicity (n [%]):** White: 146 (91%) Other: 34 (9%) **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** All patients undergoing laparoscopic evaluation of adnexal mass by gynecologic oncologist; excluded if mass above umbilicus, other indication for laparotomy, evidence of gross metastatic disease	1) Mortality: 0/160 (95% CI, 0 to 3.4%) 2) Morbidity (total all complications): 12/160 (7.5%; 95% CI, 4.3 to 12.9%) 3) Specific complications: Vascular injury: 2/160 (1.3%; 95% CI, 0.9 to 4.8%) Bleeding: 1/160 (0.6%; 0 to 3.9%) Intra-op bowel injury: 1/160 (0.6%; 0 to 3.9%) Postop bowel obstruction: 3/160 (1.9%; 0.4 to 5.7%) Postop febrile morbidity:4/160 (2.5%; 0.8 to 6.6%) 4) Rate of conversion to laparotomy: Total: 19/160 (11.9%; 95% CI, 7.8 to 18.1%) Secondary to complications: 5/160 (3.1%; 1.2 to 7.4%) 5) Final diagnoses Benign: 139 (87%; 95% CI, 84 to 90%) Borderline: 8 (5%; 2.5 to 9.9%) Ovarian cancer: 9 (5%; 2.9 to 10.6%) 4 of 9 epithelial cancers postmenopause Non-gynecological cancer: 4 (3%; 0.8 to 6.6%) 6) Other: 5 frozen sections reports changed from benign to borderline (3), malignant to benign (1), benign to cancer (1)	**Comments:** --Cases only performed by gynecological oncologists – suggests substantial prescreening in terms of likelihood of cancer, or anticipated difficulty of case **Quality assessment:** Size of population from which sample drawn: - (no data on overall referral pool) Number of cases: - (wide CIs) Patient selection: - (not much description of characteristics, likelihood of bias) Application of reference standard: + (all got pathology)

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Fanfani, Fagotti, Ercoli, et al., 2004 **#7810**	**Geographical location:** Rome, Italy **Dates:** Jan 2003-Aug 2003 **Size of population:** 100 **Single center randomized trial** **Morbidity definitions:** -Ileus -Fever (temperature ≥ 38° C on 2 consecutive measurements at least 6 hours apart) -Anemia (hemoglobin < 8 g/dl) -Bowel/bladder/ureteral injuries **Length of followup after surgery:** 30 days	**Age:** Mean (SD): Laparoscopy: 36.3 (12.1) Laparotomy: 37.5 (13.4) **Menopausal status (n [%]):** Laparoscopy: 5 (10%) postmenopausal Laparotomy: 10 (20%) postmenopausal **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** 0	**Symptomatic (n [%]):** Laparoscopy: 35 (70%) Laparotomy: 38 (76%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Randomized trial Patients excluded for BMI > 32 Cysts > 12 cm Hysterectomy required Postmenopausal and CA-125 > 35	1) Mortality: 0 2) Morbidity (total all complications): Laparoscopy: 0/50 (0%; 95% CI, 0 to 10.6%) Laparotomy: 3/50 (6.0%; 1.8 to 17.5%) 3) Specific complications: Fever: Laparoscopy: 0 Laparotomy: 2 (4%; 95% CI, 0.6 to 14.8%) Anemia: Laparoscopy : 0 Laparotomy: 1 (2%; 0 to 12.0%) 4) Rate of conversion to laparotomy: 0	**Comments:** --No malignancies --Small sample size **Quality assessment:** Size of population from which sample drawn: + Number of cases: - Patient selection: + Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Havrilesky, Peterson, Dryden, et al., 2003 **#8180**	**Geographical location:** Durham, NC **Dates:** NR **Size of population:** 396 **Single center** **Morbidity definitions:** -Estimated blood loss ≥ 500 cc -Incision disruption/infection -Urinary retention -Small bowel obstruction -Urinary tract/GI/nerve injury -Subcutaneous emphysema -Hemorrhage -Transfusion -Readmission -Re-exploration **Length of followup after surgery:** NR	**Age:** Median: 43 Range: 12-87 **Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR Post (> 55): 317 (37.2%) **Race/ethnicity (n [%]):** White: 71.2% Black: 26.2% Other: 2.4% **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** Clinical history NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: 0 (95% CI, 0 to 1.2%) 2) Morbidity (total all complications): 33/396 (8.3%; 95% CI, 6.0 to 11.6%) 3) Specific complications: Incisional disruption/infection: 7 (1.8%; 95% CI, 0.8 to 3.7%) Urinary retention: 3 (0.8%; 0.2 to 2.4%) Partial small bowel obstruction/prolonged ileus: 5 (1.3%; 0.5 to 3.1%) Urinary tract injury: 1 (0.25%; 0 to 1.6%) Bowel injury: 2 (0.5%; 0.03 to 2.0%) Nerve injury: 1 (0.25%; 0 to 1.6%) Hemorrhage: 7 (1.8%; 0.8 to 3.7%) Re-exploration: 5 (1.3%; 0.5 to 3.1%) 4) Rate of conversion to laparotomy: 25% 5) Undiagnosed cancer: 8/396 (2.0%)	**Comments:** --Presurgical history not reported --Risk of complications increased with concurrent hysterectomy --Risk of conversion increased with history of hysterectomy --8/396 (2%) had cancer, 4 (1%) had borderline disease **Quality assessment:** Size of population from which sample drawn: + Number of cases: + Patient selection: - Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Hidlebaugh, Vulga-ropulos, and Orr, 1997 **#9490**	**Geographical location:** Worcester, MA **Dates:** Jan 1988-Dec 1995 **Size of population:** 405 **Single center** **Morbidity definitions:** -Fever -Ilieus -Anemia/transfusion -Wound infection -Deep vein thrombosis -Reoperation -Readmission **Length of followup after surgery:** NR	**Age:** Range: 14-83 **Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR Post (> 55): 82 (20.2%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: 0 (95% CI, 0 to 1.38%) 2) Morbidity (total all complications): Laparoscopy: 5/199 (2.5%; 95% CI, 1.0 to 6.0%) Laparotomy: 56/206 (27.2%; 21.8 to 34.0%) 3) Specific complications: Intra-operative: Laparoscopy: 1/199 (0.5%; 95% CI, 0 to 3.2%) Laparotomy: 3/206 (1.5%; 0.3 to 4.5%) Postoperative: Laparoscopy: 2/199 (1.0%; 0.1 to 3.9%) Laparotomy: 33/206 (16.0%; 11.8 to 21.9%) Late: Laparoscopy: 2/199 (1.0%; 0.1 to 3.9%) Laparotomy: 18/206 (8.7%; 5.6 to 13.6%) Readmission: Laparoscopy: 0/199 (0%; 0 to 2.4%) Laparotomy: 2/206 (1.0%; 0.1 to 3.8%) 4) Rate of conversion to laparotomy: 10/199 (5.0%; 95% CI, 2.7 to 9.2%)	**Comments:** --Selection criteria for laparoscopy vs. laparotomy not described --Potential differences in other risk factors for complications not described --Clinical history not described **Quality assessment:** Size of population from which sample drawn: + Number of cases: + Patient selection: - Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Lok, Sahota, Rogers, et al., 2000 **#8890**	**Geographical location:** Hong Kong, China **Dates:** NR **Size of population:** 513 **Single center** **Morbidity definitions:** -Transfusion -Fever -Small bowel hernia -Pelvic hematoma -Urinary retention -Bowel/ureter/vascular injury -Reoperation **Length of followup after surgery:** NR	**Age:** Mean (SD): 35.6 (9.8) **Menopausal status (n [%]):** Pre (< 45): NR Peri (45-55): NR Post (> 55): 28 (5.5%) **Race/ethnicity (n [%]):** NR; presumably Asian **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** 389 (75.8%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** "Concurrent problems necessitating other major laparoscopic procedures" excluded	1) Mortality: 0/513 (95% CI, 0 to 1.1%) 2) Morbidity (total all complications): 68/513 (13.3%; 95% CI, 10.6 to 16.6%) 3) Specific complications: Transfusion: 0 (95% CI, 0 to 1.1%) Intra-operative: 16/513 (3.1%; 1.9 to 5.1%) Postoperative: 44 (8.6%; 6.5 to 11.4%) Febrile morbidity: 20/513 (3.9%; 2.5 to 6.0%) Reoperation: 7/513 (1.4%; 0.6 to 2.9%) 4) Rate of conversion to laparotomy: 5 (0.97%; 95% CI, 0.4 to 2.4%) 5) Undiagnosed cancer: 2/513 (0.4%)	**Comments:** --Preoperative history not described in detail --Prevalence of malignancy 0.4% --Criteria for selection for laparoscopic approach well described **Quality assessment:** Size of population from which sample drawn: + Number of cases: + Patient selection: + Application of reference standard: +
Mann and Reich, 1992 **#10330**	**Geographical location:** Kingston, PA **Dates:** NR **Size of population:** 44 **Single center** **Morbidity definitions:** Readmission **Length of followup after surgery:** NR	**Age:** Mean: 58.7 Range: 44-90 **Menopausal status (n [%]):** Post (> 55): 44 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** 7 (15.9%) **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** 37 (84.0%); proportion detected by exam vs imaging not reported **Additional data used for diagnosis:** NR	1) Mortality: 0/44 (95% CI, 0 to 11.9%) 2) Morbidity (total all complications): 2/44 (4.6%; 95% CI, 0.7 to 16.7%) 3) Specific complications: Readmission: 2/44 (4.6%; 95% CI, 0.7 to 16.7%) 4) Rate of conversion to laparotomy: 2/44 (4.6%; 95% CI, 0.7 to 16.7%) 5) Undiagnosed cancer: 1/44 (2.3%)	**Comments:** --1/44 had cancer --Ascites, effusion only exclusion **Quality assessment:** Size of population from which sample drawn: - Number of cases: - Patient selection: + Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Marana, Muzii, Catalano, et al., 2004 **#5450**	**Geographical location:** Rome, Italy **Dates:** Jul 1990 - Dec 2001 **Size of population:** 683 **Two centers (same surgeon)** **Morbidity definitions:** -Complications -Recurrence **Length of followup after surgery:** Mean 30.2 months (minimum 6 months)	**Age:** Mean: 27.6 Range: 12-39 **Menopausal status (n [%]):** Pre (< 45): 100% **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** 416 (60.9%) Chronic pain: 147 (21.5%) Dysmenorrhea: 145 (21.2%) Infertility: 66 (9.7%) Menstrual irregularity: 57 (8.3%) Abdominal swelling: 1 (0.2%) **Detected by exam (n [%]):** 267 (39.1%) (“routine”) **Detected by imaging (n [%]):** Not clear which of symptomatic ones were initially detected by imaging or exam **Combination (n [%]):** NR **Additional data used for diagnosis:** No solid components, papillation, or septae; if mass < 7 cm, ultrasound repeated in 8-12 weeks; CA-125 not done Age < 40	1) Mortality: 0/683 (95% CI, 0 to 0.82%) 2) Morbidity (total all complications): 6/683 (0.9%; 95% CI, 0.4 to 2.0%) 3) Specific complications: Retrouterine hematoma: 3/683 Febrile morbidity: 2/683 Ileus: 1/683 Umbilical hernia: 1/683 Transfusion: 1/683 4) Rate of conversion to laparotomy: 16/683 (2.3%; 95% CI, 1.3 to 3.8%) 13 patients with advanced endometriosis, 2 with large dermoids, one (0.15%) suspected malignancy (final pathology borderline) 5) 8 patients total with final path not benign – 7 borderline, 1 focal invasive endometrioid cancer	**Comments:** --Fairly complete reporting of important clinical information **--Limited to premenopausal women** --Laparotomy 1/76, laparoscopy 0/116 **Quality assessment:** Size of population from which sample drawn: - (referral base not described) Number of cases: + Patient selection: + Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Parker, Levine, Howard, et al., 1994 **#910**	**Geographical location:** Santa Monica, Irvine, and Los Angeles, CA; Louisville, KY; Rochester, NY **Dates:** NR **Size of population:** 61 **Multicenter** **Morbidity definitions:** -Complications -Conversion to laparotomy **Length of followup after surgery:** NR	**Age:** Mean: 65 Range: 47-81 **Menopausal status (n [%]):** Post (> 55): 61 (100%) postmenopausal (> 45 with 12 months of amenorrhea, or FSH > 40 mIU/mL) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Ultrasound done in all; masses < 10 cm, cystic, no irregularities, solid components, septae > 2 mm, no ascites = "presumptively benign" Post menopause = > 45 years old with at least 12 months amenorrhea, if prior hysterecotomy = FSH > 40	1) Mortality: 0/61(95% CI, 0 to 7.4%) 2) Morbidity (total all complications): 2/61 (3.3%; 95% CI, 0.4 to 12.3%) 3) Specific complications: Bladder perforation: 1/61 Sigmoid injury: 1/61 (led to laparotomy) 4) Rate of conversion to laparotomy: 3/61 (4.9%; 95% CI, 1.4 to 14.5%)	**Comments:** --Initial presentation not described --"Presumptively benign" masses form series --Only intra-operative complications reported – length of followup after surgery not reported – assume very short or not at all? **Quality assessment:** Size of population from which sample drawn: - (referral base not described) Number of cases: - Patient selection: - (not described) Application of reference standard: + (laparotomy 1/76, laparoscopy 0/116)
Parker and Proietto, 1997 **#9440**	**Geographical location:** Newcastle, Australia **Dates:** Jan 1993-Dec 1995 **Size of population:** 86 **Single center** **Morbidity definitions:** -Infection -Hernia -Pulmonary embolus -Wound hematoma or	**Age:** Mean: 34.4 Range: 12-82 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: 0/86 (95% CI, 0 to 5.4%) 2) Morbidity (total all complications): 19/86 (22.1%; 95% CI, 15.1 to 32.7%) 3) Specific complications: Wound infection: 7/86 (8.1%; 95% CI, 4.0 to 16.5%) Other infection: 5/86 (5.8%; 2.4 to 13.5%) Other wound: 2/86 (2.3%; 0.2 to 8.5%) 4) Rate of conversion to laparotomy: NA (all laparotomies) 5) Undiagnosed cancer: 1/86 (1.2%)	**Quality assessment:** Size of population from which sample drawn: Number of cases: Patient selection: Application of reference standard:

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
	other complication **Length of followup after surgery:** NR				
Sadik, Onoglu, Gokdeniz, et al., 1999 **#2880**	**Geographical location:** Matalyta, Turkey **Dates:** NR **Size of population:** 220 **Single center registry** **Morbidity definitions:** -Hospital stay -"Complications" **Length of followup after surgery:** NR	**Age:** Mean (SD): 30.0 (9.7) Range: 13-68 **Menopausal status (n [%]):** Pre (< 45): 213 (96.8%) Post (> 55): 7 (3.2%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR **Other:** 66.9% no prior surgery	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Cystic adnexal mass > 5, < 10 cm, no irregular solid parts or septae > 2mm CA-125 < 35 No ascites, matted bowel No contraindication to surgery	1) Mortality: 0/220 2) Morbidity (total all complications): 2/220 (0.9%) 3) Specific complications: Acute abdominal pain on postoperative day 5 – no cause at laparotomy Sigmoid perforation 4) Rate of conversion to laparotomy: Malignant masses "excluded from study" 5) 1 malignant dysgerminoma, 1 borderline serous cystadoma 6) 146 (67.3%) ruptured masses	**Comments:** --Data on masses converted to laparotomy not provided --Followup not described **Quality assessment:** Size of population from which sample drawn: - Number of cases: + Patient selection: - Application of reference standard: -

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Serur, Emeney, and Byrne, 2001 **#8700**	**Geographical location:** Brooklyn, NY **Dates:** Mar 1996-Nov 1998 **Size of population:** 100 (19 converted to laparotomy) **Single center** **Morbidity definitions:** -Pneumothorax -Wound infection -Fever -Enterotomy -Pneumonia **Length of followup after surgery:** Variable: 6 weeks for all All masses except patients with complex masses, ascites, and "elevated" CA-125	**Age:** Range: 17-80 (Means not given for entire group) **Menopausal status (n [%]):** Pre (< 45): 51 (51%) Post (> 55): 49 (49%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** 1/100 (1%)	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: 0/100 (0%; 95% CI, 0 to 5.44%) 2) Morbidity (total all complications): 10/100 (10%; 95% CI, 5.6 to 19.0%) 3) Specific complications: NR 4) Rate of conversion to laparotomy: 19/100 (19%; 95% CI, 12.8 to 38.6%)	**Quality assessment:** Size of population from which sample drawn: - Number of cases: + Patient selection: - Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Shalev, Eliyahu, Peleg, et al., 1994 **#10140**	**Geographical location:** Afula, Israel **Dates:** May 1988-June 1993 **Size of population:** 204; 55 underwent laparoscopy **Single center** **Morbidity definitions:** -Fever -Pain (prolonged hospitalization) **Length of followup after surgery:** NR	**Age:** NR **Menopausal status (n [%]):** Post (> 55): 204 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Inclusion: Simple or septate unilateral mass, CA-125 < 35	1) Mortality: 0/55 (95% CI, 0 to 9.6%) 2) Morbidity (total all complications): 6/55 (10.9%; 95% CI, 5.2 to 22.9%) 3) Specific complications: "Fever or pain" – not specified further 4) Rate of conversion to laparotomy: 0/55 (95% CI, 0 to 9.6%)	**Comments:** --Preoperative history not reported **Quality assessment:** Size of population from which sample drawn: - Number of cases: - Patient selection: + Application of reference standard: +
Somigliana, Ragni, Benedetti, et al., 2003 **#8140**	**Geographical location:** Milan, Italy **Dates:** Jan 2001-Dec 2002 **Size of population:** 32 **Single center** **Morbidity definitions:** Number of follicles in response to ovarian stimulation after removal of endometriotic cyst **Length of followup after surgery:** Mean 2.4 (± 1.7) years	**Age:** Mean (SD): 32.2 (3.7) **Menopausal status (n [%]):** Pre (< 45): 32 (100%) **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: NR 2) Morbidity (total all complications): Number of follicles after stimulation significantly lower in ovary where cyst removed (2.0 ± 1.5) compared to other ovary (4.2 ± 2.5) 3) Specific complications: NR 4) Rate of conversion to laparotomy: NR	**Quality assessment:** Size of population from which sample drawn: - Number of cases: - Patient selection: + Application of reference standard: -

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Tangjit-gamol, Jesadapa-trakul, Manusiri-vithaya, et al., 2004 **#1570**	**Geographical location:** Bangkok, Thailand **Dates:** Jan 1992-Dec 2002 **Size of population:** 212 **Single center** **Morbidity definitions:** Sensitivity/specificity of frozen section **Length of followup after surgery:** NR	**Age:** Mean (SD): 45.9 (17.1) Range: 13-89 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NA	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: NR 2) Morbidity (total all complications): "Defer" or greater = positive test, borderline = malignant Dis+ / Dis- / Tot T+: 84 / 8 / 92 T-: 8 / 112 / 120 Tot: 92 / 120 / 212 Value / Lower 95% CI / Upper 95% CI Se: 91.3% / 85.5% / 97.1% Sp: 93.3% / 88.9% / 97.8% PPV: 91.3% / 85.5% / 97.1% NPV: 93.3% / 88.9% / 97.8% "Defer" or greater = positive test, borderline = benign Dis+ / Dis- / Tot T+: 66 / 9 / 75 T-: 10 / 127 / 137 Tot: 76 / 136 / 212 Value / Lower 95% CI / Upper 95% CI Se: 86.8% / 79.2% / 94.4% Sp: 93.4% / 89.2% / 97.6% PPV: 88.0% / 80.6% / 95.4% NPV: 92.7% / 88.3% / 97.1% 3) Specific complications: NR 4) Rate of conversion to laparotomy: NR	**Comments:** --Presurgical history not reported **Quality assessment:** Size of population from which sample drawn: + Number of cases: + Patient selection: + Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Tarik and Fehmi, 2004 **#7770**	**Geographical location:** Ankara, Turkey **Dates:** 1996-2003 **Size of population:** 3572 (386 diagnostic laparoscopy, 1092 minor procedures) **Single center** **Morbidity definitions:** -Vascular injury -Urinary tract injury -Bowel injury -Postoperative complications **Length of followup after surgery:** NR	**Age:** Mean (SD): Diagnostic: 27.2 (4.6) Minor: 30.3 (3.2) **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: 0/3572 (95% CI, 0 to 0.13%) 2) Morbidity (total all complications): Diagnostic: 7/386 (1.8%; 95% CI, 0.8 to 3.8%) Minor: 15/1092 (1.4%; 0.8 to 2.3%) 3) Specific complications: Vascular injury: Diagnostic: 4/386 (1.0%; 95% CI, 0.3 to 2.8%) Minor: 5/1092 (0.5%; 0.2 to 1.1%) Bowel injury: Diagnostic: 2/386 (0.5%; 0.03 to 2.0%) Minor: 4) Rate of conversion to laparotomy: NR	**Comments:** --Proportion with pre-op diagnosis of adnexal mass not reported --No malignancies --Preprocedure history not reported **Quality assessment:** Size of population from which sample drawn: + Number of cases: + Patient selection: - Application of reference standard: +
van Herendael, Beretta, Slangen, et al., 1995 **#9830**	**Geographical location:** Antwerp, Belgium and Varese, Italy **Dates:** Jan 1989-Dec 1993 **Size of population:** 121 **Single center** Two centers **Morbidity definitions:** -Conversion -Anemia **Length of followup after surgery:** Median 20 months (range 1-60 months)	**Age:** Mean: 36 Range: 18-63 **Menopausal status (n [%]):** NR **Race/ethnicity (n [%]):** NR **Risk factors (n [%]):** NR **Loss to followup:** NR	**Symptomatic (n [%]):** NR **Detected by exam (n [%]):** NR **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** NR	1) Mortality: 0/121 (95% CI, 0 to 4.5%) 2) Morbidity (total all complications): 2/121 (1.7%; 95% CI, 0.1 to 6.4%) 3) Specific complications: Anemia (no transfusion): 2/121 (1.7%; 95% CI, 0.1 to 6.4%) 4) Rate of conversion to laparotomy: 3/121 (2.5%; 95% CI, 0.6 to 7.5%)	**Comments:** --Preoperative history not reported in detail **Quality assessment:** Size of population from which sample drawn: - Number of cases: - Patient selection: + Application of reference standard: +

Evidence Table 6 (continued)

Study	Study Design	Patients	Clinical Presentation	Results	Comments/Quality Scoring
Yuen, Yu, Yip, et al., 1997 **#6930**	**Geographical location:** Hong Kong **Dates:** Jul 1994-Sep 1995 **Size of population:** 110 **Single center** **Morbidity definitions:** -Complications -Cyst rupture rate **Length of followup after surgery:** 8 weeks	**Age:** Mean (SD): Laparotomy: 34.7 (8.8) Laparoscopy: 35.1 (10.3) **Menopausal status (n [%]):** Laparotomy Pre (< 45): 47 (94%) Post (> 55): 3 (6%) Laparoscopy Pre (< 45): 50 (96.2%) Post (> 55): 2 (3.8%) **Race/ethnicity (n [%]):** NR; presumably most Asian **Risk factors (n [%]):** NR **Loss to followup:** 4/110 (3.6%)	**Symptomatic (n [%]):** Laparotomy: 32 (64%) Laparoscopy: 28 (54%) **Detected by exam (n [%]):** Asymptomatic: Laparotomy 18 (36%) Laparoscopy 24 (46.1%) **Detected by imaging (n [%]):** NR **Combination (n [%]):** NR **Additional data used for diagnosis:** Cystic masses with no irregular solid parts, thick septae, ascites; except dermoid	1) Mortality: Laparotomy: 0/50 (95% CI, 0 to 10.6%) Laparoscopy: 0/52 (0 to 10.2%) 2) Morbidity (total all complications): Total: Laparotomy: 14/50 (28%; 95% CI, 18.5 to 43.1%) Laparoscopy: 4/52 (9.6%; 4.2 to 21.8%) Intraoperative Laparotomy: 1/50 (2.0%; 0 to 12.0%) Laparoscopy: 1/52 (1.9%; 0 to 11.6%) Postoperative: Laparotomy: 13/50 (26%; 16.8 to 41.0%) Laparoscopy 4/52 (7.8%; 2.9 to 19.3%) 3) Specific complications: Laparotomy: Bladder injury: 1/50 Febrile morbidity: 10/50 UTI: 5/50 Urinary retention: 4/50 Wound infection: 1/50 Laparoscopy: Inf epigastric artery injury: 1/52 Febrile morbidity: 3/52 UTI: 2/52 4) Rate of conversion to laparotomy: 0/52 5) Cyst rupture rate Lapartomy: 9/30 (30%; 95% CI, 18.4 to 50.3%) Laparoscopy: 9/33 (27.3%; 16.5 to 46.4%)	**Comments:** -- Well-defined complications **Quality assessment:** Size of population from which sample drawn: - Number of cases: + (a priori sample size calculation) Patient selection: + Application of reference standard: +

Evidence Table 7: *Question 7: What are the estimated trade-offs resulting from various strategies for evaluation of the adnexal mass?*

Study	Study Design	Study Outcomes	Sources for Model Probabilities	Sources for Model Outcomes	Results	Comments
Schapira, Matchar, and Young, 1993 **#4870**	**Type of model:** Decision tree **Population modeled (age, range):** Cohort of healthy 40-year old women. Life expectancy based on average life expectancy in the US for a 40-year old woman. **Strategies compared**: No screening to screening using CA 125 and TVS in combination	Life-expectancy No costs included	No data used for transition probabilities. One time screen assumed. Used NCI data to determine prevalence of disease in 40 year old population. Adjusted for % that would present with symptoms. National Halothane study used for estimate of laparotomy mortality **Simplifying assumptions:** Assume that survival time for early disease detected by screening or clinical symptoms is equivalent Morbidity and mortality rates for diagnostic laparotomy are the same for pts with and without the disease No benefit from identifying benign disease One-time screen	DEALE used to determine life expectancy. 1988 NCHS figures used to estimate average life expectancy. Note European citations 31 and 32 for life expectancy for those with early and late stage cancer.	1) CA125+TVS – 40.192 years No Screen – 40.191 years 2) No screening preferred if post-operative mortality rate > 7.32% or specificity of the test is 98.35% 3) Findings similar although LE gains not as great for women aged 65+. Specificity of test ≥ 99.25% in order for screening to be favored.	Progression of ovarian cancer assumed to proceed in stepwise fashion through stages

Evidence Table 7 (continued)

Study	Study Design	Study Outcomes	Sources for Model Probabilities	Sources for Model Outcomes	Results	Comments
Skates and Singer, 1991 **#7000**	**Type of model:** Stochastic simulation model **Population modeled: (age, range):** Age 50 to 75 **Strategies compared:** CA-125 screening	Years of life saved (undiscounted)	Assumes o log-normal distribution for each stage o correlation between duration of adjacent stages is high, lower for stages far apart o coefficient of variation constant across all stages Estimates and ranges for duration of stage obtained from 2 gynecologic oncologists Stage I: 9 months Stage II: 4.5 months Stage III: 12 months Stage IV: 3 months Range corresponded to CV of approximately 50% Stage at clinical detection is independent of duration of disease; "major determinant of detection being the size and spread of the tumor and not it's rate of growth" CA-125 produced from "tumor inception" Screen detected cases have identical survival within each	Massachusetts General Hospital survival data for cancer Life tables for competing cause of death	Based on "best" estimates of mean duration and variation of duration, 3.4 ± 1 year of life saved per case by annual screening; range of 1-5 years Overall increase in life expectancy (cases and noncases) not reported Results sensitive to assumptions about duration of early stage cancer, frequency of screening	--Model assumes stepwise progression through disease stages ("all tumors are assumed to pass through all four stages if there is no intervention" --Source of estimates for natural history parameters obtained from only 2 gynecologic oncologists (not discussed in article, only evident from reference 11 (personal communications) --Assumption regarding duration of disease and likelihood of detection doesn't necessarily reflect biology of disease --Model output of stage distribution in unscreened population not reported.

Evidence Table 7 (continued)

Study	Study Design	Study Outcomes	Sources for Model Probabilities	Sources for Model Outcomes	Results	Comments
			stage as clinically detected cases Survival independent of age			
Tengs, Winer, Paddock, et al., 1998 **#6010**	**Type of model:** Markov model **Population modeled (age, range):** 30-year old woman testing for BRCA 1 and 2 **Strategies compared**: Testing for BRCA 1 and 2 with the following surgical options: Do nothing Mastectomy Mastectomy and oophorectomy Oophorectomy	Life years saved quality adjusted life years saved No costs included	Test accuracy provided by survey of companies marketing the tests SEER data for incidence – adjusted to account for the fact that data does not distinguish between those with genetic risk and those without. Literature used to estimate 92% risk reduction for breast cancer. Cancer experts for effectiveness of oophorectomy w/wout mastectomy **Simplifying assumptions:** "Operative mortality not included in the model as it would not have…an appreciable effect."	SEER data and literature (Rubin, NEJM, 1996) used for ovarian cancer 5 year survival NCHS for non-breast ovarian cancer survival Unclear where data for utilities came from	1) Immediate mastectomy + oophorectomy offers greatest gains in survival when measured using LE Testing offers no benefit. 2) Optimal intervention depends on pre-test probability of carrying mutation. 3) When QALYs incorporated, depends on the test characteristics. If test is perfectly sens and spec then maximizes QALYs. If not perfectly sens and spec, then depends again on pre-test probability of mutation.	--Only considers prophylactic oophorectomy in setting of positive BRCA1/2 test; testing for early stage disease not considered --Natural history of ovarian cancer not explicitly modeled

Evidence Table 7 (continued)

Study	Study Design	Study Outcomes	Sources for Model Probabilities	Sources for Model Outcomes	Results	Comments
Urban, Drescher, Etzioni and Colby 1997 **#6140**	**Type of model:** Stochastic simulation model **Population modeled: (age, range):** Age 50 to 80 **Strategies compared:** TVS and CA-125	Years of life saved	Age and stage based on SEER data Length of Stage 1 assumed to be independent of stage of disease at clinical diagnosis; Model assumes disease stages that correspond to FIGO staging with durations that are distributed log-normally with geometric means of 9, 4.5 12 and 3 months respectively, Stage 2 – ½ of Stage I; Stage 3, 1.333 times the length of Stage I; Stage 4, .333 times the length of Stage 1. Used data from Skates and Singer to model natural history 0.001 probability of death each time laparotomy is performed.	Costs are based on a survey of labs, hospital clinics, and physician offices in Seattle, WA Note: used a 5% discount rate for the analysis	Multimodal strategy using CA 125 with a threshold for positivity of either elevation above 35U/ml or doubling since the previous screen, followed by TVS only if CA 125 is positive was found to be efficient. (Cost/year of life saved is $64,000) Effectiveness and cost-effectiveness sensitive to assumptions about the behavior of early stage disease	--Model assumes stepwise progression through disease stages; assumption based on opinion of clinicians (Skates and Singer 1991, above)

Abbreviations used in the Evidence Tables

2D	Two-dimensional
3D	Three-dimensional
AFP	Alpha-fetoprotein
AHRQ	Agency for Healthcare Research and Quality
AUC	Area under the curve
BME	Bimanual examination
BMI	Body mass index
CA-19-9	Cancer antigen 19-9
CA-72-4	Cancer antigen 72-4
CA-125	Cancer antigen 125
CEA	Carcinoembryonic antigen
CI	Confidence interval
CPP	Chronic pelvic pain
CT	Computed tomography
F-FDG	18-Fluorodeoxyglucose
FNA	Fine needle aspiration
FSH	Follicle-stimulating hormone
GI	Gastrointestinal
hCG	Human chorionic gonadotropin
ICD-9	*International Classification of Diseases, Ninth Revision*
LDH	Lactate dehydrogenase
LMP	Low malignant potential
MRI	Magnetic resonance imaging
NIS	Nationwide Inpatient Sample
NA	Not applicable
NPV	Negative predictive value
NR	Not reported
OR	Odds ratio
PE	Pelvic examination
PET	Positron emission tomography
PI	Pulsatility index
PID	Pelvic inflammatory disease
PPS	Papillary projection score
PPV	Positive predictive value
PSV	Peak systolic velocity
RI	Resistance index
RMI	Risk of Malignancy Index
ROC	Receiver operating characteristic
SD	Standard deviation
Se	Sensitivity
SEM	Standard error of the mean
Sp	Specificity
TAG-72	Tumor-associated glycoprotein 72
TAMXV	Time-averaged maximum velocity

TATI	Tumor-associated trypsin inhibitor
TVUS	Transvaginal ultrasound
US	Ultrasound
UTI	Urinary tract infection

References Cited in the Evidence Tables

Adonakis GL, Paraskevaidis E, Tsiga S, et al. A combined approach for the early detection of ovarian cancer in asymptomatic women. Eur J Obstet Gynecol Reprod Biol 1996;65(2):221-5.

Alcazar JL, Castillo G. Comparison of 2-dimensional and 3-dimensional power-Doppler imaging in complex adnexal masses for the prediction of ovarian cancer. Am J Obstet Gynecol 2005;192(3):807-12.

Alcazar JL, Errasti T, Zornoza A, et al. Transvaginal color Doppler ultrasonography and CA-125 in suspicious adnexal masses. Int J Gynaecol Obstet 1999;66(3):255-61.

Alcazar JL, Galan MJ, Garcia-Manero M, et al. Three-dimensional sonographic morphologic assessment in complex adnexal masses: preliminary experience. J Ultrasound Med 2003;22(3):249-54.

Alcazar JL, Lopez-Garcia G. Transvaginal color Doppler assessment of venous flow in adnexal masses. Ultrasound Obstet Gynecol 2001;17(5):434-8.

Alcazar JL, Merce LT, Laparte C, et al. A new scoring system to differentiate benign from malignant adnexal masses. Am J Obstet Gynecol 2003;188(3):685-92.

Anandakumar C, Chew S , Wong YC , et al. Role of transvaginal ultrasound color flow imaging and Doppler waveform analysis in differentiating between benign and malignant ovarian tumors. Ultrasound Obstet Gynecol 1996;7:280-4.

Andersen ES, Knudsen A, Rix P, et al. Risk of malignancy index in the preoperative evaluation of patients with adnexal masses. Gynecol Oncol 2003;90(1):109-12.

Andolf E, Jorgensen C, Astedt B. Ultrasound examination for detection of ovarian carcinoma in risk groups. Obstet Gynecol 1990;75(1):106-9.

Antonic J, Rakar S. Colour and pulsed Doppler US and tumour marker CA 125 in differentiation between benign and malignant ovarian masses. Anticancer Res 1995;15:1527-32.

Asif N, Sattar A, Dawood MM, et al. Pre-operative evaluation of ovarian mass: risk of malignancy index. J Coll Physicians Surg Pak 2004;14(3):128-31.

Aslam N, Banerjee S, Carr JV, et al. Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer. Obstet Gynecol 2000;96(1):75-80.

Aslam N, Tailor A, Lawton F, et al. Prospective evaluation of three different models for the pre-operative diagnosis of ovarian cancer. BJOG 2000;107(11):1347-53.

Balbi GC, Musone R, Menditto A, et al. Women with a pelvic mass: indicators of malignancy. Eur J Gynaecol Oncol 2001;22(6):459-62.

Benjapibal M, Sunsaneevitayakul P, Boriboonhirunsarn D, et al. Color Doppler ultrasonography for prediction of malignant ovarian tumors. J Med Assoc Thai 2002;85(6):709-15.

Benjapibal M, Sunsaneevitayakul P, Phatihattakorn C, et al. Sonographic morphological pattern in the pre-operative prediction of ovarian masses. J Med Assoc Thai 2003;86(4):332-7.

Berlanda N, Ferrari MM, Mezzopane R, et al. Impact of a multiparameter, ultrasound-based triage on surgical management of adnexal masses. Ultrasound Obstet Gynecol 2002;20(2):181-5.

Biagiotti R, Desii C, Vanzi E, et al. Predicting ovarian malignancy: application of artificial neural networks to transvaginal and color Doppler flow US. Radiology 1999;210(2):399-403.

Bromley B, Goodman H, Benacerraf BR. Comparison between sonographic morphology and Doppler waveform for the diagnosis of ovarian malignancy. Obstet Gynecol 1994;83(3):434-7.

Brown DL, Doubilet PM, Miller FH, et al. Benign and malignant ovarian masses: selection of the most discriminating gray-scale and Doppler sonographic features. Radiology 1998;208(1):103-10.

Buckshee K, Temsu I, Bhatla N, et al. Pelvic examination, transvaginal ultrasound and transvaginal color Doppler sonography as predictors of ovarian cancer. Int J Gynaecol Obstet 1998;61(1):51-7.

Buist MR, Golding RP, Burger CW, et al. Comparative evaluation of diagnostic methods in ovarian carcinoma with emphasis on CT and MRI. Gynecol Oncol 1994;52(2):191-8.

Buy JN, Ghossain MA, Hugol D, et al. Characterization of adnexal masses: combination of color Doppler and conventional sonography compared with spectral Doppler analysis alone and conventional sonography alone. AJR Am J Roentgenol 1996;166(2):385-93.

Canis M, Mage G, Pouly JL, et al. Laparoscopic diagnosis of adnexal cystic masses: a 12-year experience with long-term follow-up. Obstet Gynecol 1994;83(5 Pt 1):707-12.

Canis M, Mashiach R, Wattiez A, et al. Frozen section in laparoscopic management of macroscopically suspicious ovarian masses. J Am Assoc Gynecol Laparosc 2004;11(3):365-9.

Canis M, Pouly JL, Wattiez A, et al. Laparoscopic management of adnexal masses suspicious at ultrasound. Obstet Gynecol 1997;89(5 Pt 1):679-83.

Carley ME, Klingele CJ, Gebhart JB, et al. Laparoscopy versus laparotomy in the management of benign unilateral adnexal masses. J Am Assoc Gynecol Laparosc 2002;9(3):321-6.

Carter JR, Lau M, Fowler JM, et al. Blood flow characteristics of ovarian tumors: implications for ovarian cancer screening. Am J Obstet Gynecol 1995;172(3):901-7.

Carter PG, Iles RK, Neven P, et al. The measurement of urinary beta core fragment in conjunction with serum CA125 does not aid the differentiation of malignant from benign pelvic masses. Gynecol Oncol 1993;51(3):368-71.

Caruso A, Caforio L, Testa AC, et al. Transvaginal color Doppler ultrasonography in the presurgical characterization of adnexal masses. Gynecol Oncol 1996;63(2):184-91.

Castillo G, Alcazar JL, Jurado M. Natural history of sonographically detected simple unilocular adnexal cysts in asymptomatic postmenopausal women. Gynecol Oncol 2004;92(3):965-9.

Chalas E, Welshinger M, Engellener W, et al. The clinical significance of thrombocytosis in women presenting with a pelvic mass. Am J Obstet Gynecol 1992;166(3):974-7.

Chapron C, Dubuisson JB, Capella-Allouc S. Salpingo-oophorectomy for adnexal masses. Place and results for operative laparoscopy. Eur J Obstet Gynecol Reprod Biol 1997;73(1):43-8.

Chen DX, Schwartz PE, Li FQ. Saliva and serum CA 125 assays for detecting malignant ovarian tumors. Obstet Gynecol 1990;75(4):701-4.

Chen DX, Schwartz PE, Li XG, et al. Evaluation of CA 125 levels in differentiating malignant from benign tumors in patients with pelvic masses. Obstet Gynecol 1988;72(1):23-7.

Chi DS, Abu-Rustum NR, Sonoda Y, et al. Ten-year experience with laparoscopy on a gynecologic oncology service: analysis of risk factors for complications and conversion to laparotomy. Am J Obstet Gynecol 2004;191(4):1138-45.

Childers JM, Nasseri A, Surwit EA. Laparoscopic management of suspicious adnexal masses. Am J Obstet Gynecol 1996;175(6):1451-7; discussion 1457-9.

Chou CY, Chang CH, Yao BL , et al. Color Doppler ultrasonography and serum CA 125 in the differentiation of benign and malignant ovarian tumors. J Clin Ultrasound 1994;22:491-6.

Cohen LS, Escobar PF, Scharm C, et al. Three-dimensional power Doppler ultrasound improves the diagnostic accuracy for ovarian cancer prediction. Gynecol Oncol 2001;82(1):40-8.

Davies AP, Jacobs I, Woolas R, et al. The adnexal mass: benign or malignant? Evaluation of a risk of malignancy index. Br J Obstet Gynaecol 1993;100(10):927-31.

Deckardt R, Saks M, Graeff H. Comparison of minimally invasive surgery and laparotomy in the treatment of adnexal masses. J Am Assoc Gynecol Laparosc 1994;1(4 Pt 1):333-8.

DePriest PD, Gallion HH, Pavlik EJ, et al. Transvaginal sonography as a screening method for the detection of early ovarian cancer. Gynecol Oncol 1997;65(3):408-14.

DePriest PD, Shenson D, Fried A, et al. A morphology index based on sonographic findings in ovarian cancer. Gynecol Oncol 1993;51(1):7-11.

DePriest PD, van Nagell JR Jr, Gallion HH, et al. Ovarian cancer screening in asymptomatic postmenopausal women. Gynecol Oncol 1993;51(2):205-9.

DePriest PD, Varner E, Powell J. The efficacy of a sonographic morphology index in identifying ovarian cancer: a multi-institutional investigation. Gynecol Oncol 1994;55:174-8.

Dottino PR, Levine DA, Ripley DL, et al. Laparoscopic management of adnexal masses in premenopausal and postmenopausal women. Obstet Gynecol 1999;93(2):223-8.

Dowd JR, Quinn MA, Rome R, et al. Women with a pelvic mass--when to perform an ultrasound. Aust N Z J Obstet Gynaecol 1993;33(4):404-7.

Einhorn N, Bast RC Jr, Knapp RC, et al. Preoperative evaluation of serum CA 125 levels in patients with primary epithelial ovarian cancer. Obstet Gynecol 1986;67(3):414-6.

Ekerhovd E, Wienerroith H, Staudach A, et al. Preoperative assessment of unilocular adnexal cysts by transvaginal ultrasonography: a comparison between ultrasonographic morphologic imaging and histopathologic diagnosis. Am J Obstet Gynecol. 2001;184(2):48-54.

Fanfani F, Fagotti A, Ercoli A, et al. A prospective randomized study of laparoscopy and minilaparotomy in the management of benign adnexal masses. Hum Reprod 2004;19(10):2367-71.

Fenchel S, Grab D, Nuessle K, et al. Asymptomatic adnexal masses: correlation of FDG PET and histopathologic findings. Radiology 2002;223(3):780-8.

Ferdeghini M, Gadducci A, Prontera C, et al. Serum soluble interleukin-2 receptor assay in epithelial ovarian cancer. Tumour Biol 1993;14(5):303-9.

Ferrazzi E, Zanetta G, Dordoni D, et al. Transvaginal ultrasonographic characterization of ovarian masses: comparison of five scoring systems in a multicenter study.[erratum appears in Ultrasound Obstet Gynecol 1998 Jan;11(1):v Note: Lissoni G[corrected to Lissoni AA]]. Ultrasound Obstet Gynecol 1997;10(3):192-7.

Finkler NJ, Benacerraf B, Lavin PT, et al. Comparison of serum CA 125, clinical impression, and ultrasound in the preoperative evaluation of ovarian masses. Obstet Gynecol 1988;72(4):659-64.

Fleischer AC, Cullinan JA, Jones HW 3rd, et al. Correlation of histomorphology of ovarian masses with color Doppler sonography. Ultrasound Med Biol 1996;22(5):555-9.

Fleischer AC, Rodgers WH, Kepple DM, et al. Color Doppler sonography of benign and malignant ovarian masses. Radiographics 1992;12(5):879-85.

Franchi M, Beretta P, Ghezzi F, et al. Diagnosis of pelvic masses with transabdominal color Doppler, CA 125 and ultrasonography. Acta Obstet Gynecol Scand 1995;74(9):734-9.

Gadducci A, Baicchi U, Marrai R, et al. Preoperative evaluation of D-dimer and CA 125 levels in differentiating benign from malignant ovarian masses. Gynecol Oncol 1996;60(2):197-202.

Gadducci A, Capriello P, Bartolini T, et al. The association of ultrasonography and CA-125 test in the preoperative evaluation of ovarian carcinoma. Eur J Gynaecol Oncol 1988;9(5):373-6.

Gadducci A, Ferdeghini M, Prontera C, et al. The concomitant determination of different tumor markers in patients with epithelial ovarian cancer and benign ovarian masses: relevance for differential diagnosis. Gynecol Oncol 1992;44(2):147-54.

Gadducci A, Ferdeghini M, Rispoli G, et al. Comparison of tumor-associated trypsin inhibitor (TATI) with CA125 as a marker for diagnosis and monitoring of epithelial ovarian cancer. Scand J Clin Lab Invest Suppl 1991;207(19-24).

Goldstein SR, Subramanyam B, Snyder JR, et al. The postmenopausal cystic adnexal mass: the potential role of ultrasound in conservative management. Obstet Gynecol 1989;73(1):8-10.

Grab D, Flock F, Stohr I, et al. Classification of asymptomatic adnexal masses by ultrasound, magnetic resonance imaging, and positron emission tomography. Gynecol Oncol 2000;77(3):454-9.

Granberg S, Norstrom A, Wikland M. Tumors in the lower pelvis as imaged by vaginal sonography. Gynecol Oncol 1990;37(2):224-9.

Granberg S, Norstrom A, Wikland M. Comparison ofendovaginal ultrasound and cytological evaluation of cystic ovarian tumors. J Ultrasound Med 1991;10:9-14.

Grover SR, Quinn MA. Is there any value in bimanual pelvic examination as a screening test. Med J Aust 1995;162(8):408-10.

Guerriero S, Ajossa S, Garau N, et al. Ultrasonography and color Doppler-based triage for adnexal masses to provide the most appropriate surgical approach. Am J Obstet Gynecol 2005;192(2):401-6.

Guerriero S, Ajossa S, Risalvato A, et al. Diagnosis of adnexal malignancies by using color Doppler energy imaging as a secondary test in persistent masses. Ultrasound Obstet Gynecol 1998;11(4):277-82.

Guerriero S, Alcazar JL, Coccia ME, et al. Complex pelvic mass as a target of evaluation of vessel distribution by color Doppler sonography for the diagnosis of adnexal malignancies: results of a multicenter European study. J Ultrasound Med 2002;21(10):1105-11.

Hata K, Hata T, Kitao M. Intratumoral peak systolic velocity as a new possible predictor for detection of adnexal malignancy. Am J Obstet Gynecol 1995;172:1496-500.

Hata K, Hata T, Manabe A, et al. A critical evaluation of transvaginal Doppler studies, transvaginal sonography, magnetic resonance imaging, and CA 125 in detecting ovarian cancer. Obstet Gynecol 1992;80(6):922-6.

Havrilesky LJ, Peterson BL, Dryden DK, et al. Predictors of clinical outcomes in the laparoscopic management of adnexal masses. Obstet Gynecol 2003;102(2):243-51.

Herrmann UJ Jr, Locher GW, Goldhirsch A. Sonographic patterns of ovarian tumors: prediction of malignancy. Obstet Gynecol 1987;69(5):777-81.

Hidlebaugh DA, Vulgaropulos S, Orr RK. Treating adnexal masses. Operative laparoscopy vs. laparotomy. J Reprod Med 1997;42(9):551-8.

Hillaby K, Aslam N, Salim R, et al. The value of detection of normal ovarian tissue (the 'ovarian crescent sign') in the differential diagnosis of adnexal masses. Ultrasound Obstet Gynecol 2004;23(1):63-7.

Hogdall EV, Hogdall CK, Tingulstad S, et al. Predictive values of serum tumour markers tetranectin, OVX1, CASA and CA125 in patients with a pelvic mass. Int J Cancer 2000;89(6):519-23.

Hricak H, Chen M, Coakley FV, et al. Complex adnexal masses: detection and characterization with MR imaging--multivariate analysis. Radiology 2000;214(1):39-46.

Huber S, Medl M, Baumann L, et al. Value of ultrasound and magnetic resonance imaging in the preoperative evaluation of suspected ovarian masses. Anticancer Res 2002;22(4):2501-7.

Hurteau JA, Woolas RP, Jacobs IJ, et al. Soluble interleukin-2 receptor alpha is elevated in sera of patients with benign ovarian neoplasms and epithelial ovarian cancer. Cancer 1995;76(9):1615-20.

Inoue M, Fujita M, Nakazawa A, et al. Sialyl-Tn, sialyl-Lewis Xi, CA 19-9, CA 125, carcinoembryonic antigen, and tissue polypeptide antigen in differentiating ovarian cancer from benign tumors. Obstet Gynecol 1992;79(3):434-40.

Itakura T, Kikkawa F, Kajiyama H, et al. Doppler flow and arterial location in ovarian tumors. Int J Gynaecol Obstet 2003;83(3):277-83.

Jacobs I, Oram D, Fairbanks J, et al. A risk of malignancy index incorporating CA 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer. Br J Obstet Gynaecol 1990;97(10):922-9.

Jacobs I, Stabile I, Bridges J, et al. Multimodal approach to screening for ovarian cancer. Lancet 1988;1(8580):268-71.

Jain KA. Prospective evaluation of adnexal masses with endovaginal gray-scale and duplex and color Doppler US: correlation with pathologic findings. Radiology 1994;191(1):63-7.

Jain KA, Friedman DL, Pettinger TW, et al. Adnexal masses: comparison of specificity of endovaginal US and pelvic MR imaging. Radiology 1993;186(3):697-704.

Juhasz B, Kurjak A, Lampe L, et al. Tissue characterization by transvaginal colour doppler for the evaluation of gynaecological tumours. Acta Med Hung 1990;47:149-56.

Kawahara K, Yoshida Y, Kurokawa T, et al. Evaluation of positron emission tomography with tracer 18-fluorodeoxyglucose in addition to magnetic resonance imaging in the diagnosis of ovarian cancer in selected women after ultrasonography. J Comput Assist Tomogr 2004;28(4):505-16.

Kawai M, Kikkawa F, Ishikawa H. Differential diagnosis of ovarian tumors by transvaginal color -pulse Doppler sonography. Gynecol Oncol 1994;54:209-14.

Komatsu T, Konishi I, Mandai M, et al. Adnexal masses: transvaginal US and gadolinium-enhanced MR imaging assessment of intratumoral structure. Radiology 1996;198(1):109-15.

Kurjak A, Kupesic S. Three dimensional ultrasound and power doppler in assessment of uterine and ovarian angiogenesis: a prospective study. Croat Med J 1999;40(3):413-20.

Kurjak A, Kupesic S, Sparac V, et al. Three-dimensional ultrasonographic and power Doppler characterization of ovarian lesions. Ultrasound Obstet Gynecol 2000;16(4):365-71.

Kurjak A, Predanic M. New scoring system for prediction of ovarian malignancy based on transvaginal color Doppler sonography. J Ultrasound Med 1992;11(12):631-8.

Kurjak A, Schulman H, Sosic A, et al. Transvaginal ultrasound, color flow, and Doppler waveform of the postmenopausal adnexal mass. Obstet Gynecol 1992;80(6):917-21.

Kurjak A, Shalan H, Kupesic S, et al. An attempt to screen asymptomatic women for ovarian and endometrial cancer with transvaginal color and pulsed Doppler sonography. J Ultrasound Med 1994;13(4):295-301.

Kurjak A, Zalud I, Alfirevic Z. Evaluation of adnexal masses with transvaginal color ultrasound. J Ultrasound Med 1991;10(6):295-7.

Kurtz AB, Tsimikas JV, Tempany CM, et al. Diagnosis and staging of ovarian cancer: comparative values of Doppler and conventional US, CT, and MR imaging correlated with surgery and histopathologic analysis--report of the Radiology Diagnostic Oncology Group. Radiology 1999;212(1):19-27.

Leeners B, Schild RL, Funk A, et al. Colour Doppler sonography improves the pre-operative diagnosis of ovarian tumours made using conventional transvaginal sonography. Eur J Obstet Gynecol Reprod Biol 1996;64(1):79-85.

Lerner JP, Timor-Tritsch IE, Federman A, et al. Transvaginal ultrasonographic characterization of ovarian masses with an improved, weighted scoring system. Am J Obstet Gynecol 1994;170(1 Pt 1):81-5.

Levine D, Gosink BB, Wolf SI, et al. Simple adnexal cysts: the natural history in postmenopausal women.[see comment]. Radiology 1992;184(3):653-9.

Lin JY, Angel C, DuBeshter B, et al. Diagnoses after laparotomy for a mass in the pelvic area in women. Surg Gynecol Obstet 1993;176(4):333-8.

Lin P, Wu C, Lee C, et al. Color Doppler Ultrasound in the Assessment of Ovarian Neoplasms. J Med Ultrasound 1993;1(4):172-6.

Lok IH, Sahota DS, Rogers MS, et al. Complications of laparoscopic surgery for benign ovarian cysts. J Am Assoc Gynecol Laparosc 2000;7(4):529-34.

Lu C, Van Gestel T, Suykens JA, et al. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. Artif Intell Med 2003;28(3):281-306.

Luxman D, Bergman A, Sagi J, et al. The postmenopausal adnexal mass: correlation between ultrasonic and pathologic findings. Obstet Gynecol 1991;77(5):726-8.

Ma S, Shen K, Lang J. A risk of malignancy index in preoperative diagnosis of ovarian cancer. Chin Med J 2003;116(3):396-9.

Maggino T, Gadducci A, D'Addario V, et al. Prospective multicenter study on CA 125 in postmenopausal pelvic masses. Gynecol Oncol 1994;54(2):117-23.

Malkasian GD Jr, Knapp RC, Lavin PT, et al. Preoperative evaluation of serum CA 125 levels in premenopausal and postmenopausal patients with pelvic masses: discrimination of benign from malignant disease. Am J Obstet Gynecol 1988;159(2):341-6.

Maly Z, Riss P, Deutinger J. Localization of blood vessels and qualitative assessment of blood flow in ovarian tumors. Obstet Gynecol 1995;85(1):33-6.

Mancuso A, De Vivo A, Triolo O, et al. The role of transvaginal ultrasonography and serum CA 125 assay combined with age and hormonal state in the differential diagnosis of pelvic masses. Eur J Gynaecol Oncol 2004;25(2):207-10.

Manjunath AP, Pratapkumar, Sujatha K, et al. Comparison of three risk of malignancy indices in evaluation of pelvic masses. Gynecol Oncol 2001;81(2):225-9.

Mann WJ, Reich H. Laparoscopic adnexectomy in postmenopausal women. J Reprod Med 1992;37(3):254-6.

Marana R, Muzii L, Catalano GF, et al. Laparoscopic excision of adnexal masses. J Am Assoc Gynecol Laparosc 2004;11(2):162-6.

Marchetti M, Zambon A, Lamaina V, et al. Ultrasound as a possible screening method in ovarian cancer.[erratum appears in Eur J Gynaecol Oncol. 2002;23(3):172 Note: Marchioro S [corrected to Marchiori S]]. Eur J Gynaecol Oncol 2002;23(2):123-6.

Marret H, Sauget S, Giraudeau B, et al. Contrast-enhanced sonography helps in discrimination of benign from malignant adnexal masses. J Ultrasound Med. 2004;23(12):1629-39; quiz 1641-42.

Matthes AC, Moreira de Andrade JM , Bighetti S. Selection of criteria for the treatment of ovarian cysts on the bases of ultrasound and cytology. Gynecol Obstet Invest 1996;42:244-8.

McIntosh MW, Drescher C, Karlan B, et al. Combining CA125 and SMR serum markers for diagnosis and early detection of ovarian cancer. Gynecol Oncol 2004;95:9-15.

Medl M, Kulenkampff KJ, Stiskal M, et al. Magnetic resonance imaging in the preoperative evaluation of suspected ovarian masses. Anticancer Res 1995;15(3):1123-5.

Menon U, Talaat A, Rosenthal AN, et al. Performance of ultrasound as a second line test to serum CA125 in ovarian cancer screening. BJOG 2000;107(2):165-9.

Merce LT, Caballero RA, Barco MJ, et al. B-mode, utero-ovarian and intratumoural transvaginal colour Doppler ultrasonography for differential diagnosis of ovarian tumours. Eur J Obstet Gynecol Reprod Biol 1998;76(1):97-107.

Modesitt SC, Pavlik EJ, Ueland FR, et al. Risk of malignancy in unilocular ovarian cystic tumors less than 10 centimeters in diameter. Obstet Gynecol 2003;102(3):594-9.

Mol BW, Boll D, De Kanter M, et al. Distinguishing the benign and malignant adnexal mass: an external validation of prognostic models. Gynecol Oncol 2001;80(2):162-7.

Morgante G, la Marca A, Ditto A, et al. Comparison of two malignancy risk indices based on serum CA125, ultrasound score and menopausal status in the diagnosis of ovarian masses. Br J Obstet Gynaecol 1999;106(6):524-7.

O'Connell GJ, Ryan E, Murphy KJ, et al. Predictive value of CA 125 for ovarian carcinoma in patients presenting with pelvic masses. Obstet Gynecol 1987;70(6):930-2.

Obeidat BR, Amarin ZO, Latimer JA, et al. Risk of malignancy index in the preoperative evaluation of pelvic masses. Int J Gynaecol Obstet 2004;85(3):255-8.

Ong S, Duffy T, Murphy J. Transabdominal ultrasound and its correlation with clinical findings in gynaecology. Ir J Med Sci 1996;165(4):268-70.

Onsrud M, Shabana A, Austgulen R. Soluble tumor necrosis factor receptors and CA 125 in serum as markers for epithelial ovarian cancer. Tumour Biol 1996;17(2):90-6.

Padilla LA, Radosevich DM, Milad MP. Accuracy of the pelvic examination in detecting adnexal masses. Obstet Gynecol 2000;96(4):593-8.

Padilla LA, Radosevich DM, Milad MP. Limitations of the pelvic examination for evaluation of the female pelvic organs. Int J Gynaecol Obstet 2005;88(1):84-8.

Padungsutt P, Thirapagawong C, Senapad S, et al. Accuracy of tissue polypeptide specific antigen (TPS) in the diagnosis of ovarian malignancy. Anticancer Res 2000;20(2B):1291-5.

Parker J, Proietto AM. Surgical treatment of adnexal cysts by laparotomy. Int J Gynaecol Obstet 1997;59(3):257-8.

Parker WH, Levine RL, Howard FM, et al. A multicenter study of laparoscopic management of selected cystic adnexal masses in postmenopausal women. J Am Coll Surg 1994;179(6):733-7.

Patsner B, Mann WJ. The value of preoperative serum CA 125 levels in patients with a pelvic mass. Am J Obstet Gynecol 1988;159(4):873-6.

Prompeler HJ, Madjar H, Sauerbrei W. Classification of adnexal tumors by transvaginal color Doppler. Gynecol Oncol 1996;61(3):354-63.

Pyrgiotis E, Salamalekis E, Loghis C, et al. CA 125 in preoperative evaluation of pelvic masses. Eur J Gynaecol Oncol 1993;14(4):279-82.

Rehn M, Lohmann K, Rempen A. Transvaginal ultrasonography of pelvic masses: evaluation of B-mode technique and Doppler ultrasonography. Am J Obstet Gynecol 1996;175(1):97-104.

Reles A, Wein U, Lichtenegger W. Transvaginal color Doppler sonography and conventional sonography in the preoperative assessment of adnexal masses. J Clin Ultrasound 1997;25(5):217-25.

Reuter M, Steffens J. , Schuppler U. Critical evaluation of the specificity of MRI and TVUs for differentiation of malignant from benign adnexal lesions. Eur Radiol 1998;8(39-44).

Roman LD, Muderspach LI, Burnett AF, et al. Carcinoembryonic antigen in women with isolated pelvic masses. Clinical utility? J Reprod Med 1998;43(5):403-7.

Roman LD, Muderspach LI, Stein SM, et al. Pelvic examination, tumor marker level, and gray-scale and Doppler sonography in the prediction of pelvic cancer. Obstet Gynecol 1997;89(4):493-500.

Sadik S, Onoglu AS, Gokdeniz R, et al. Laparoscopic management of selected adnexal masses. J Am Assoc Gynecol Laparosc 1999;6(3):313-6.

Salem S, White LM, Lai J. Doppler sonography of adnexal masses: the predictive value of the pulsatility index in benign and malignant disease. AJR Am J Roentgenol 1994;163(5):1147-50.

Sassone AM, Timor-Tritsch IE, Artner A, et al. Transvaginal sonographic characterization of ovarian disease: evaluation of a new scoring system to predict ovarian malignancy. Obstet Gynecol 1991;78(1):70-6.

Sawicki W, Spiewankiewicz B, Cendrowski K, et al. Preoperative discrimination between malignant and benign adnexal masses with transvaginal ultrasonography and colour blood flow imaging. Eur J Gynaecol Oncol 2001;22(2):137-42.

Schapira MM, Matchar DB, Young MJ. The effectiveness of ovarian cancer screening. A decision analysis model. Ann Intern Med 1993;118(11):838-43.

Schelling M, Braun M, Kuhn W, et al. Combined transvaginal B-mode and color Doppler sonography for differential diagnosis of ovarian tumors: results of a multivariate logistic regression analysis. Gynecol Oncol 2000;77(1):78-86.

Schincaglia P, Brondelli L, Cicognani A, et al. A feasibility study of ovarian cancer screening: does fine-needle aspiration improve ultrasound specificity? Tumori 1994;80(3):181-7.

Schneider VL, Schneider A, Reed KL, et al. Comparison of Doppler with two-dimensional sonography and CA 125 for prediction of malignancy of pelvic masses. Obstet Gynecol 1993;81(6):983-8.

Schutter EM, Davelaar EM, van Kamp GJ, et al. The differential diagnostic potential of a panel of tumor markers (CA 125, CA 15-3, and CA 72-4 antigens) in patients with a pelvic mass. Am J Obstet Gynecol 2002;187(2):385-92.

Schutter EM, Kenemans P, Sohn C, et al. Diagnostic value of pelvic examination, ultrasound, and serum CA 125 in postmenopausal women with a pelvic mass. An international multicenter study. Cancer 1994;74(4):1398-406.

Schutter EM, Sohn C, Kristen P, et al. Estimation of probability of malignancy using a logistic model combining physical examination, ultrasound, serum CA 125, and serum CA 72-4 in postmenopausal women with a pelvic mass: an international multicenter study. Gynecol Oncol 1998;69(1):56-63.

Scoutt LM, McCarthy SM, Lange R, et al. MR evaluation of clinically suspected adnexal masses. J Comput Assist Tomogr 1994;18(4):609-18.

Sengoku K, Satoh T, Saitoh S, et al. Evaluation of transvaginal color Doppler sonography, transvaginal sonography and CA 125 for prediction of ovarian malignancy. Int J Gynaecol Obstet 1994;46(1):39-43.

Serur E, Emeney PL, Byrne DW. Laparoscopic management of adnexal masses. J Soc Laparoendosc Surg 2001;5(2):143-51.

Shabana A, Onsrud M. Tissue polypeptide-specific antigen and CA 125 as serum tumor markers in ovarian carcinoma. Tumour Biol 1994;15(6):361-7.

Shalev E, Eliyahu S, Peleg D, et al. Laparoscopic management of adnexal cystic masses in postmenopausal women. Obstet Gynecol 1994;83(4):594-6.

Shen-Gunther J, Mannel RS. Ascites as a predictor of ovarian malignancy. Gynecol Oncol 2002;87(1):77-83.

Siegel BA, Dehdashti F, Mutch DG, et al. Evaluation of 111In-DTPA-folate as a receptor-targeted diagnostic agent for ovarian cancer: initial clinical results. J Nucl Med 2003;44(5):700-7.

Skates SJ, Singer DE. Quantifying the potential benefit of CA 125 screening for ovarian cancer. J Clin Epidemiol 1991;44(4-5):365-80.

Smikle CB, Lunt CC, Hankins GD. Clinical predictors in the evaluation of a pelvic mass. Mil Med 1995;160(5):233-5.

Sohaib SA, Mills TD, Sahdev A, et al. The role of magnetic resonance imaging and ultrasound in patients with adnexal masses. Clin Radiol 2005;60(3):340-8.

Somigliana E, Ragni G, Benedetti F, et al. Does laparoscopic excision of endometriotic ovarian cysts significantly affect ovarian reserve? Insights from IVF cycles. Hum Reprod 2003;18(11):2450-3.

Soper JT, Hunter VJ, Daly L, et al. Preoperative serum tumor-associated antigen levels in women with pelvic masses. Obstet Gynecol 1990;75(2):249-54.

Stein SM, Laifer-Narin S, Johnson MB, et al. Differentiation of benign and malignant adnexal masses: relative value of gray-scale, color Doppler, and spectral Doppler sonography. AJR Am J Roentgenol 1995;164(2):381-6.

Strigini FA, Gadducci A, Del Bravo B, et al. Differential diagnosis of adnexal masses with transvaginal sonography, color flow imaging, and serum CA 125 assay in pre- and postmenopausal women. Gynecol Oncol 1996;61(1):68-72.

Szpurek D, Moszyniki R, Sajdak S. Clinical value of the ultrasound Doppler index in determination of ovarian tumor malignancy. Eur J Gynaecol Oncol 2004;25(4):442-4.

Tailor A, Bourne TH, Campbell S, et al. Results from an ultrasound-based familial ovarian cancer screening clinic: a 10-year observational study. Ultrasound Obstet Gynecol 2003;21(4):378-85.

Tailor A, Jurkovic D, Bourne TH, et al. Sonographic prediction of malignancy in adnexal masses using an artificial neural network. Br J Obstet Gynaecol 1999;106(1):21-30.

Takac I. Analysis of blood flow in adnexal tumors by using color Doppler imaging and pulsed spectral analysis. Ultrasound Med Biol 1998;24(8):1137-42.

Tangjitgamol S, Jesadapatrakul S, Manusirivithaya S, et al. Accuracy of frozen section in diagnosis of ovarian mass. Int J Gynecol Cancer 2004;14(2):212-9.

Tanir HM, Ozalp S, Yalcin OT, et al. Preoperative serum vascular endothelial growth factor (VEGF) in ovarian masses. Eur J Gynaecol Oncol 2003;24(3-4):271-4.

Tarik A, Fehmi C. Complications of gynaecological laparoscopy--a retrospective analysis of 3572 cases from a single institute. J Obstet Gynaecol 2004;24(7):813-6.

Tay SK, Chua EK. Correlation of serum, urinary and salivary CA 125 levels in patients with adnexal masses. Ann Acad Med Singapore 1994;23(3):311-4.

Tekay A, Jouppila P. Validity of pulsatility and resistance indices in classification of adnexal tumors with transvaginal color doppler ultrasound. Ultrasound Obstet Gynecol 1992;2:338-44.

Tengs TO, Winer EP, Paddock S, et al. Testing for the BRCA1 and BRCA2 breast-ovarian cancer susceptibility genes: a decision analysis. Med Decis Making 1998;18(4):365-75.

Tepper R, Lerner-Geva L, Altaras MM, et al. Transvaginal color flow imaging in the diagnosis of ovarian tumors. J Ultrasound Med 1995;14(10):731-4.

Tian J, Zhang J, Jiao L, et al. A prospective study of Tc-99m MIBI in the differential diagnosis of pelvic masses in female patients. Clin Nucl Med 2000;25(8):614-8.

Timmerman D, Bourne TH, Tailor A, et al. A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model. Am J Obstet Gynecol 1999;181(1):57-65.

Timmerman D, Verrelst H, Bourne TH, et al. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. Ultrasound Obstet Gynecol 1999;13(1):17-25.

Timor-Tritsch LE, Lerner J, Monteagudo A, et al. Transvaginal ultrasonographic characterization of ovarian masses by means of color flow-directed Doppler measurements and a morphologic scoring system. Am J Obstet Gynecol 1993;168:909-13.

Tingulstad S, Hagen B, Skjeldestad FE, et al. The risk-of-malignancy index to evaluate potential ovarian cancers in local hospitals. Obstet Gynecol 1999;93(3):448-52.

Tingulstad S, Hagen B, Skjeldestad FE, et al. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. Br J Obstet Gynaecol 1996;103(8):826-31.

Torres JC, Derchain SF, Faundes A, et al. Risk-of-malignancy index in preoperative evaluation of clinically restricted ovarian cancer. Rev Paul Med 2002;120(3):72-6.

Troiano RN, Quedens-Case C, Taylor KJ. Correlation of findings on transvaginal sonography with serum CA 125 levels. AJR Am J Roentgenol 1997;168(6):1587-90.

Twickler DM, Forte TB, Santos-Ramos R, et al. The Ovarian Tumor Index predicts risk for malignancy. Cancer 1999;86(11):2280-90.

Urban N, Drescher C, Etzioni R, et al. Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening. Control Clin Trials 1997;18(3):251-70.

Valentin L. Gray scale sonography, subjective evaluation of the color Doppler image and measurement of blood flow velocity for distinguishing benign and malignant tumors of suspected adnexal origin. Eur J Obstet Gynecol Reprod Biol 1997;72(1):63-72.

Valentin L. Pattern recognition of pelvic masses by gray-scale ultrasound imaging: the contribution of Doppler ultrasound. Ultrasound Obstet Gynecol 1999;14(5):338-47.

Valentin L. Comparison of Lerner score, Doppler ultrasound examination, and their combination for discrimination between benign and malignant adnexal masses. Ultrasound Obstet Gynecol 2000;15(2):143-7.

Valentin L, Akrawi D. The natural history of adnexal cysts incidentally detected at transvaginal ultrasound examination in postmenopausal women. Ultrasound Obstet Gynecol 2002;20(2):174-80.

Valentin L, Hagen B, Tingulstad S, et al. Comparison of 'pattern recognition' and logistic regression models for discrimination between benign and malignant pelvic masses: a prospective cross validation. Ultrasound Obstet Gynecol 2001;18(4):357-65.

van Herendael B, Beretta P, Slangen T, et al. Management of adnexal masses by operative laparoscopy. J Am Assoc Gynecol Laparosc 1995;2(3):273-7.

van Nagell JR Jr, DePriest PD, Reedy MB, et al. The efficacy of transvaginal sonographic screening in asymptomatic women at risk for ovarian cancer. Gynecol Oncol 2000;77(3):350-6.

Vasilev SA, Schlaerth JB, Campeau J, et al. Serum CA 125 levels in preoperative evaluation of pelvic masses. Obstet Gynecol 1988;71(5):751-6.

Vuento MH, Pirhonen JP, Makinen JI, et al. Evaluation of ovarian findings in asymptomatic postmenopausal women with color Doppler ultrasound. Cancer 1995;76(7):1214-8.

Wakahara F, Kikkawa F, Nawa A, et al. Diagnostic efficacy of tumor markers, sonography, and intraoperative frozen section for ovarian tumors. Gynecol Obstet Invest 2001;52(3):147-52.

Weiner Z, Thaler I, Beck D, et al. Differentiating malignant from benign ovarian tumors with transvaginal color flow imaging. Obstet Gynecol 1992;79(2):159-62.

Woolas RP, Conaway MR, Xu F, et al. Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. Gynecol Oncol 1995;59(1):111-6.

Wu CC, Lee CN, Chen T, et al. Factors contributing to the accuracy in diagnosing ovarian malignancy by color Doppler ultrasound. Obstet Gynecol 1994;84:605-8.

Yamashita Y, Hatanaka Y, Torashima M, et al. Characterization of sonographically indeterminate ovarian tumors with MR imaging. A logistic regression analysis. Acta Radiol 1997;38(4 Pt 1):572-7.

Yamashita Y, Torashima M, Hatanaka Y, et al. Adnexal masses: accuracy of characterization with transvaginal US and precontrast and postcontrast MR imaging. Radiology 1995;194(2):557-65.

Yuen PM, Yu KM, Yip SK, et al. A randomized prospective study of laparoscopy and laparotomy in the management of benign ovarian masses. Am J Obstet Gynecol 1997;177(1):109-14.

Zanetta G, Vergani P, Lissoni A. Color Doppler ultrasound in the preoperative assessment of adnexal masses. Acta Obstet Gynecol Scand 1994;73:637-41.

Zhang Z, Barnhill SD, Zhang H, et al. Combination of multiple serum markers using an artificial neural network to improve specificity in discriminating malignant from benign pelvic masses. Gynecol Oncol 1999;73(1):56-61.

Zimmer Y, Tepper R, Akselrod S. An automatic approach for morphological analysis and malignancy evaluation of ovarian masses using B-scans. Ultrasound Med Biol 2003;29(11):1561-70.

Appendix E: Peer Reviewers

The Duke Evidence-based Practice Center is grateful to the following peer reviewers who read and commented on a draft version of this report:

Susan M. Ascher, MD; Department of Radiology, Georgetown University Hospital; Washington, DC

Andrew Berchuck, MD; Division of Gynecologic Oncology, Duke University Medical Center; Durham, NC

Michael L. Berman; Division of Gynecologic Oncology, UCI Medical Center; Orange, CA

Christie R. Eheman, MD; Centers for Disease Control and Prevention; Atlanta, GA

Barbara Goff, MD; University of Washington School of Medicine; Seattle, WA

Walter Kinney, MD; Department of Obstetrics and Gynecology, The Permanente Medical Group; Sacramento, CA

Ann Kolker; Ovarian Cancer National Alliance; Washington, DC

Herschel Lawson, MD; Center for Disease Control and Prevention; Atlanta, GA

Saralyn Mark, MD; Department of Health and Human Services; Washington, DC

Susan Meikle, MD, MSPH; Agency for Healthcare Research and Quality; Rockville, MD

Valerie McGuire, PhD; Department of Health Research and Policy, Stanford University; Stanford, CA

Edward E. Partridge, MD; Department of Obstetrics and Gynecology, University of Alabama; Birmingham, AL

Mona Saraiya, MD, MPH; Centers for Disease Control and Prevention; Atlanta, GA

George F. Sawaya, MD; Department of Obstetrics and Gynecology, UCSF; San Francisco, CA

Howard T. Sharp, MD; University of Utah Hospitals and Clinics; Salt Lake City, UT

Edward L. Trimble, MD, MPH; National Cancer Institute; Rockville, MD

John R. van Nagell Jr., MD; Department of Obstetrics and Gynecology, University of Kentucky Medical Center; Lexington, Kentucky

Nominations for peer reviewers were solicited from several sources, including the project's technical expert panel and interested federal agencies. The list of nominees was vetted and approved by the Agency for Healthcare Research and Quality (AHRQ).