On December 6, 1999, under Public Law 106-129, the Agency for Health Care Policy Research (AHCPR) was reauthorized and renamed the Agency for Healthcare Research and Quality (AHRQ). The law authorizes AHRQ to continue its research on the cost, quality, and outcomes of health care, and expands its role to improve patient safety and address medical errors.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Systematic Evidence Review
Number 5

Newborn Hearing Screening

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
2101 East Jefferson Street
Rockville, MD 20852
http://www.ahrq.gov

Contract No. 290-97-0018
Technical Support of the U.S. Preventive Services Task Force

Prepared by:
Oregon Health & Science University
Evidence-based Practice Center, Portland, Oregon

Mark Helfand, MD, MS
Diane C. Thompson, MS
Robert Davis, MD
Heather McPhillips, MD
Charles J. Homer, MD, MPH
Tracy L. Lieu, MD, MPH

AHRQ Publication No. 02-S001
October 2001

Preface

The Agency for Healthcare Research and Quality (AHRQ) sponsors the development of Systematic Evidence Reviews (SERs) through its Evidence-based Practice Program. With guidance from the third U.S. Preventive Services Task Force* (USPSTF) and input from Federal partners and primary care specialty societies, two Evidence-based Practice Centers—one at the Oregon Health & Science University and the other at Research Triangle Institute-University of North Carolina—systematically review the evidence of the effectiveness of a wide range of clinical preventive services, including screening, counseling, immunizations, and chemoprevention, in the primary care setting. The SERs—comprehensive reviews of the scientific evidence on the effectiveness of particular clinical preventive services—serve as the foundation for the recommendations of the third USPSTF, which provide age- and risk-factor-specific recommendations for the delivery of these services in the primary care setting. Details of the process of identifying and evaluating relevant scientific evidence are described in the "Methods" chapter of each SER.

The SERs document the evidence regarding the benefits, limitations, and cost-effectiveness of a broad range of clinical preventive services and will help to further awareness, delivery, and coverage of preventive care as an integral part of quality primary health care.

AHRQ also disseminates the SERs on the AHRQ Web site (http://www.ahrq.gov/uspstfix.htm) and disseminates summaries of the evidence (summaries of the SERs) and recommendations of the third USPSTF in print and on the Web. These are available through the AHRQ Web site (http://www.ahrgq.gov/uspstfix.htm), through the National Guideline Clearinghouse (http://www.ncg.gov), and in print through the AHRQ Publications Clearinghouse (1-800-358-9295).

We welcome written comments on this SER. Comments may be sent to: Director, Center for Practice and Technology Assessment, Agency for Healthcare Research and Quality, 6010 Executive Blvd., Suite 300, Rockville, MD 20852.

John M. Eisenberg, M.D.
Director
Agency for Healthcare Research and Quality

Robert Graham, M.D.
Director, Center for Practice and Technology Assessment
Agency for Healthcare Research and Quality

* The USPSTF is an independent panel of experts in primary care and prevention first convened by the U.S. Public Health Service in 1984. The USPSTF systematically reviews the evidence on the effectiveness of providing clinical preventive services—including screening, counseling, immunization, and chemoprevention—in the primary care setting. AHRQ convened the third USPSTF in November 1998 to update existing Task Force recommendations and to address new topics.

Structured Abstract

Context. Each year approximately 5000 infants are born in the United States with moderate to profound, bilateral sensorineural hearing loss (SNHL). Universal newborn hearing screening (UNHS) has been proposed as a means to speed diagnosis and treatment, and thereby improve language outcomes in these children.

Objective. To identify strengths, weaknesses, and gaps in the evidence supporting UNHS and to compare the additional benefits and harms of UNHS with those of selective screening of high-risk newborns.

Data Sources. A keyword search of MEDLINE, CINAHL, and PsycINFO databases for relevant papers published from 1994 to August 2001, using terms for hearing disorders, infant or newborn, screening, and relevant treatments. We contacted experts and reviewed reference lists to identify additional articles, including those published before 1994.

Study Selection. We included controlled and observational studies of (1) the accuracy, yield, or harms of screening using otoacoustic emissions (OAEs), auditory brainstem response (ABR), or both in the general newborn population or (2) the effects of screening or of early identification and treatment on language outcomes. Nineteen articles, including 1 controlled trial, met these inclusion criteria.

Data Extraction. Data on population, test performance, outcomes, and methodological quality were extracted using prespecified criteria developed by the US Preventive Services Task Force. We queried authors when information needed to assess study quality was missing.

Data Synthesis. Good quality studies show from 2041 to 2794 low-risk, and 86 to 208 high-risk, newborns were screened to find 1 case of moderate to profound SNHL. The best estimate of positive predictive value is 6.7%. Six percent to 15% of infants who fail the screening tests are subsequently diagnosed with bilateral SNHL. In a trial of UNHS versus clinical screening at 8 months of age, UNHS increased the proportion of infants with moderate to severe hearing loss diagnosed by 10 months of age (57% vs 14%), but did not reduce the rate of diagnosis after 18 months of age. No good-quality controlled study has compared UNHS to selective screening of high-risk newborns. In fair- to poor-quality cohort studies, intervention before 6 months of age was associated with improved language and communication skills by 2 to 5 years of age. These studies had unclear criteria for selecting subjects, and none compared an inception cohort of low-risk newborns identified by screening to those identified in usual care, making it impossible to exclude selection bias as an explanation for the results. In a mathematical model based on the literature review, we estimated that extending screening to low-risk infants would detect 1 additional case before 10 months for every 1441 low-risk infants screened, and result in treatment before 10 months of 1 additional case for every 2401 low-risk infants screened. With UNHS, 254 newborns would be referred for audiological evaluation because of false-positive second-stage screening test results, versus 48 for selective screening.

Conclusions. Modern screening tests for hearing impairment can improve identification of newborns with SNHL, but the efficacy of UNHS to improve long-term language outcomes remains uncertain.

Suggested Citation:
Helfand M, Thompson DC, Davis R, et al. Newborn Hearing Screening. Systematic Evidence Review Number 5 (Contract 290-97-0018 to the Oregon Health & Science University Evidence–based Practice Center, Portland, Oregon). AHRQ Publication No. 02-S001. Rockville, MD: Agency for Healthcare Research and Quality. October 2001.

CONTENTS

1. INTRODUCTION

Each year in the United States, approximately 5000 infants are born with moderate, severe, or profound bilateral sensorineural hearing loss (SNHL). Hearing loss in childhood is associated with poor language development in early childhood and with lower educational achievement and employment opportunities later in life.[1-5]

For children without risk factors, hearing loss frequently escapes detection until the age when hearing children normally begin to talk (at 9 months of age or older).[6-10] Current theory views auditory stimuli during the first 6 months of life as critical to development of speech and language skills.[11-13] Advocates of universal newborn hearing screening (UNHS) believe that earlier application of available therapies, such as speech and language therapy, amplification, and family support, could reduce or eliminate the gap in language skills between deaf and hearing children.[14, 15] Screening in the hospital prior to discharge also presents an opportunity to provide services to children with limited access to routine medical care, and is one way to deliver state-of-the-art services to disadvantaged and under-served populations.

The incidence of SNHL varies with race, birthweight, and other risk factors. Among infants in a neonatal intensive care unit (NICU), the risk of moderate to severe SNHL is 10 to 20 times higher than the general population.[16] The Joint Committee on Infant Hearing high risk guidelines specify admission of 2 or more days to an NICU or one of 4 other risk factors (Table 1).[17] From 10% to 30% of newborns meet these criteria, which can identify 50% to 75% of all cases of moderate to profound bilateral hearing loss.[18]

In 1995 the US Preventive Services Task Force (USPSTF) found insufficient evidence to recommend UNHS.[19] They argued that, among low-risk infants, the prevalence of hearing impairment was very low, and substantial numbers of infants would be misclassified. They found that evidence for the efficacy of early intervention for patients diagnosed by screening was incomplete, but endorsed selective screening of high-risk newborns based on the higher prevalence of hearing loss in this group.

Since 1995, many health care professionals and federal health care agencies have advocated for UNHS, which is now mandated by law in 32 states.[6, 17, 20] Is this widespread support for UNHS now justified? To update the USPSTF recommendations, we critically reviewed recent evidence to identify strengths, weaknesses, and gaps in the evidence supporting UNHS.

A. Epidemiology and Burden of Illness

Estimates of the prevalence of moderate to profound bilateral hearing loss vary, depending on the criteria used to define the different degrees of hearing loss and the characteristics of the studied population.[21] When a criterion of $\geq$ 40 dB HL is used to diagnose moderate hearing loss, the prevalence of permanent, congenital bilateral hearing loss ranges from 1 in 900 to 1 in 2,500 newborns.[18, 22-26] From one-third to two-thirds of these have moderate bilateral hearing loss, while the remainder have severe or profound hearing loss.

The prevalence of congenital hearing loss also depends on race, birthweight, and other risk factors. In a well-done, population-based survey of 3-year olds in Atlanta, Georgia, the overall prevalence of congenital bilateral hearing loss was 1 in 2400 in whites and 1 in 1350 in blacks.[26] Among low birthweight infants (<2500 grams), 1 in 500 black children and 1 in 714 white children had bilateral hearing loss. Among infants admitted to an intensive care unit, the risk of moderate to severe SNHL is 10 to 20 times as high as the healthy nursery population. In the Rhode Island Hearing Assessment Project, for example, 1 in 76 newborns had hearing impairment in the intensive care population, versus 1 in 775 in the well-baby nursery.[16] The prevalence is lower still in the subset of well infants who have no family history of hearing impairment.

It is clear that the diagnosis of congenital hearing impairment is often delayed. A British study found that 19% of children with congenital hearing loss were identified by the age of 6 months and 39% by their first year.[7] It is thought that, in the United Kingdom, about a quarter of children remain undiagnosed until after 42 months of age.[8] In 1993, a National Institutes of Health panel estimated that the average age of identification in the US was about 3 years.[6] In a

survey of 331 parents from 35 states,[10] performed in the early 1990s, the median age at diagnosis was 12.5 months for all children with severe and profound congenital hearing loss and 17.2 months for all children with mild and moderate losses. For children with no known risk factors for hearing loss, the median age of diagnosis was 13 months for those with severe to profound hearing loss and 22 months for children with mild-moderate hearing loss. In a more recent survey conducted by the Center for Assessment and Demographic Studies at Gallaudet University, the mean age of diagnosis of deaf children was 14.5 months.[9]

On average, children with hearing loss have delayed development in vocabulary, grammar, conversation, and reading.[1-5] This delay is measurable before 3 years of age[27] and has consequences throughout life. The 1997-98 Gallaudet Research Institute's annual survey of deaf and hard-of-hearing children and youth nationwide (n=30,198) found that half of the students had communication limitations, and almost half had at least one cognitive, behavioral, or social limitation.[28] The average deaf student graduates from high school with language and academic achievement levels below those of the average fourth-grade student with normal hearing.[29, 30] Average reading scores for hard-of-hearing students graduating from high school are at the fifth-grade level.[29, 31] The lag in reading performance has remained virtually unchanged since it was first carefully measured in the early 1960s.[32, 33]

The contribution of delayed identification to this lag in communication is difficult to assess. One-third to one-half of infants with congenital hearing impairment have other developmental disabilities, most often mental retardation and sequelae of prematurity, that contribute to poor language development. The degree of parental involvement and the availability of special educational facilities, which may also contribute to language outcomes, are not controlled for in most frequently cited studies of the development of language skills.

In fact, the natural history of congenital hearing loss in the subgroup that is the primary target of universal screening has not been studied carefully. The studies that establish that hearing impairment has serious consequences do not focus on the subset of children who have no risk factors for congenital hearing loss. This group, which is comprised of children who did not

require NICU admission and do not have other risk factors for hearing impairment, may have a lower prevalence of coexisting disabilities than do hearing impaired children in general. The frequency, severity, and, especially, the duration of language delay throughout childhood and into adult life in this group is not known.

Screening also detects mild, fluctuating, and unilateral SNHL, as well as conductive hearing loss. These conditions are common in school-age children and clearly delay speech and language acquisition and harm school performance.[34-36] Very little information is available about how often mild impairment detected at birth proves to be temporary or indistinguishable from normal in followup testing. Many studies address the consequences of mild and unilateral hearing loss that is acquired after infancy, but the frequency and significance of these conditions detected at birth is not clear.

B. Health Care Interventions

Screening Tests

Two types of tests are commonly used to screen for congenital hearing loss: otoacoustic emissions (OAEs) and auditory brainstem response (ABR). OAE testing evaluates the integrity of the inner ear (cochlea). In response to noise, vibrations of the hair cells in a healthy inner ear generate electrical responses, known as otoacoustic emissions. The absence of OAEs indicates that the inner ear is not responding appropriately to sound. Transient evoked otoacoustic emissions (TEOAEs) are generated in response to wide-band clicks, while distortion product otoacoustic emissions (DPOAE) are a response to tones. Both stimuli are presented via a light-weight ear canal probe. A microphone picks up the signal, and multiple responses are averaged to get a specific repeatable waveform.

The ABR is an electrophysiological response generated in the brainstem in response to auditory signals and composed of either clicks or tones. The stimulus is delivered via earphones or an inserted ear probe, and scalp electrodes pick up the signal. ABR evaluates the integrity of the peripheral auditory system and the auditory nerve pathways up to the brainstem and is able to

identify infants with normal cochlear function but abnormal eighth-nerve function (auditory neuropathy).

Typically, screening programs use a 2-stage screening approach (either OAE repeated twice, OAE followed by ABR, or ABR repeated twice). Criteria for defining a "pass" or "fail" on the initial screening test vary widely. In a survey of 25 programs, for example, 21 different "pass" criteria were being used.[37]

ABR and OAE have limitations that affect their accuracy in certain patients. Both require a sleeping or quiet child. Middle-ear effusion or debris in the external canal can compromise the accuracy of these tests. OAE and ABR test the peripheral auditory system and eighth nerve pathway to the brainstem, respectively. They are not designed to identify infants with central hearing deficits. Therefore, infants with risk factors for central hearing deficits, particularly those who have congenital Cytomegalovirus infection or prolonged severe hypoxia at birth, may pass their newborn hearing screens with either OAE or ABR, but develop profound hearing loss in early infancy.[38, 39]

The newer generation of automated screeners are easy to use and do not require highly trained staff. However, equipping hospitals with equipment and sufficient staff can be costly, the staff must be trained to understand the limitations of the techniques, and ongoing quality control is essential to achieve accurate, consistent test results. The importance of technique is illustrated by the results of multicenter studies of universal screening, in which the rates of false positive and technically inadequate examinations varied ten-fold among sites.[40]

Confirmation of the Final Diagnosis

A behavioral test, such as visual reinforcement audiometry (VRA), is the appropriate gold standard determination for permanent hearing impairment.[41] Because VRA cannot be performed reliably before 8 to 9 months of age,[42, 43] studies of screening use an intermediate diagnostic standard to follow up results of screening tests. The intermediate diagnostic standard

usually consists of diagnostic ABR testing or other electrophysiological testing, along with an otolaryngological examination, [43] These assessments have traditionally been performed after 6 months of age, but in some programs are done as early as 2 months of age. The accuracy of the intermediate diagnostic standard depends on the age at which it is performed. The interobserver variability of these assessments in infants has not been evaluated in high-quality studies, and the frequency of over- and under-diagnosis is not known.

Further Management

A well-defined, well-established treatment protocol of proven effectiveness is often cited as a prerequisite to adoption of a new screening program.[17] Management of infants who have hearing impairment is multifaceted, reflecting the complex pathology of poor language development as well as uncertainty about the benefits and risks of alternative treatment plans. Programs for infants and children who have impaired hearing usually include a component of family therapy, intended to prepare parents and siblings to care for and communicate with a deaf child. Hearing aids, cochlear implants, auditory trainers, and communication systems (speech, cued speech, sign language) are also used.[5, 44]

Different experts advocate substantially different approaches based on competing theories of language acquisition and communication. This variation is reflected in practice, even among programs in states or hospitals with established screening programs. In a survey of 500 programs in 15 states, 388 programs indicated that they provide early intervention services.[44] The programs varied in their fundamental approach to instruction. About half of the programs offered an Auditory-Verbal approach (emphasizing "the acquisition of spoken language through specific auditory training techniques"), while the other half offered an Auditory-Oral approach (stressing "the use of residual hearing, speech, and oral language development"). Half of the programs offered American Sign Language, and 57% offered an English-based sign system with simultaneous speech. Other high-variation components of care were the methods used to assess

progress in children's skills; availability of medical specialists (56%), psychologists (70%), and deaf adults (62%) as consultants or instructors; and the use of parental support groups (51%).

The reason for this variation is that the evidence basis for many components of the intervention is weak.[5] The efficacy of early intervention programs for hearing impaired children, and the individual components of these programs, have not been established in randomized trials or in population-based cohort studies.

Practices regarding the timing of interventions also vary. Variation in the timing of treatment could limit the effectiveness of newborn screening. Current theory argues for initiation of treatment before 6 months of age.[11-13] In one retrospective, population-based study of hearing-impaired children born in Trent, United Kingdom, during the late 1980s, the median age at referral, confirmation of the impairment, prescription of the hearing aid, and fitting of the hearing aid were, respectively, 10.4 months, 18.1 months, 24.4 months and 26.3 months.[18] On average, 7.7 months passed between the time a child was referred for audiological examination and the time a final diagnosis was made. It is not clear whether this discrepancy was due to administrative delays, parental decisions, or the need for observation over time to confirm the suspected diagnosis. A lag time of 6.3 months was observed between the time of diagnosis and the time that a hearing aid was prescribed. Survey results from the United States showed a great deal of variation in the lag time between diagnostic confirmation and onset of interventions.[9] The average lag times ranged from 8.1 months for hearing aids to 11.2 months for sign language. Lag times were longer for children with milder losses, for those who had deaf parents, and for sign language, Hispanic and black children. However, these factors account for only a small portion of the variation in lag times and do not explain the high lag times observed in children without these factors.

C. Analytic Framework and Key Questions

The key assumptions underlying UNHS are that 1) diagnosis and, therefore, treatment are delayed until after age 1 or 2 in many of these children; 2) delays result in anxiety and stress on the family and the infant; 3) selective screening of high–risk newborns misses as many as half the cases of moderate to profound hearing impairment; 4) universal screening would reduce the frequency of delay; and 5) early treatment results in better language function in the preschool period, leading to better educational, occupational, and social function later in life.

As noted earlier, solid data support the first assumption, and there is a plausible case, based on anecdotal evidence, for the second. Recent studies of UNHS address the last three assumptions. Figure 1, the analytic framework for this systematic review, depicts assumptions that underlie UNHS in a diagram of the populations, interventions, intermediate outcomes, and health outcomes affected. Available studies of the effects of early intervention focus on the intermediate outcome of language and communication skills in preschool children. The framework notes that, in the long-term, the health outcomes of primary interest might include mental health, psychosocial and cognitive function, and school and occupational performance throughout life. The purpose of our systematic review was to assess whether there was sufficient evidence about each link in the analytic framework to estimate the net effect of universal screening indirectly from separate bodies of evidence about universal screening programs and about alternatives, such as high-risk screening.

Figure 2 lists key questions related to each link in the analytic framework. Our search and review of the literature was organized around these questions. To assess how accurately screening can diagnose congenital hearing impairment, we asked questions about the sensitivity and specificity of the OAE and ABR tests and about the yield of screening in actual practice, taking compliance with screening and followup into account (Arrow 1).

Randomized trials or well-done cohort studies would provide the best evidence linking UNHS to improved language outcomes in infants *who would not be diagnosed early in a high-risk screening program* (Arrow 3). No such studies have been published. Instead, inferences

about the effect of early treatment rely on observational studies that examine the association between birth at a hospital where UNHS was practiced, or the age at diagnosis of hearing loss, and language ability in children of preschool age (Arrow 4). The fundamental question is, does treatment prior to 6 months lead to improved language and communication in infants who would not be diagnosed that early in a selective, high-risk screening program? This additional benefit due to screening must outweigh the harms of screening in children who do not have hearing impairment (Arrow 2) and the potential harms, if any, due to early diagnosis and treatment in children who have hearing impairment (Arrow 5).

2. METHODS

A. Literature Search

Before we arrived at the final key questions listed in Figure 2, we considered a longer list of questions that potentially could have been answered in a literature review. This list included more specific questions about the consequences of screening and questions about benefits of treatment other than language skills and school performance. We conducted a preliminary MEDLINE search to eliminate questions for which we were unable to find evidence. For example, we did not find studies that evaluated the significance of poor communication skills on social functioning and occupational opportunities, so our outcome measures are limited to language and communication skills during early childhood in the final list of key questions.

To find relevant articles on screening for hearing impairment, we searched the MEDLINE, CINAHL, and PsycINFO databases for papers published from 1994 to September 2000. (See Appendix 1 for the MEDLINE search string.) We also used reference lists of comprehensive review articles[8, 21, 45-53] and expert recommendations to locate additional articles published after 1994. We relied on the 1995 USPSTF's review[19] and suggestions of experts and peer reviewers to identify important articles published in 1994 or earlier. Searches were updated monthly through August 2001.

Searches of the electronic databases returned 864 abstracts. Two investigators reviewed each abstract to determine whether to obtain the full text of the article. Disagreement was resolved by discussion. Abstracts were included if both reviewers agreed that the topic was relevant to one of the key questions and the article contained data. Abstracts were also included if there was not sufficient information to classify the article. The full-text versions of 177 articles from the searches and about 30 articles from other sources were obtained and examined by 2 reviewers for inclusion in evidence tables.

Full-text articles were included in the systematic review if they were 1) controlled trials, 2) reports on the accuracy, yield, or harms of screening from state-based, population-based studies, or hospital-based UNHS programs using ABR or OAE technology in the general newborn population, or 3) reports of the effects of screening, early identification and treatment, or any type of language outcomes. For the last group, we excluded uncontrolled case series and case reports. [54-62] We excluded studies in which screening was done with physical examination or with tests other than ABR or OAE.[37, 63-77] We also included studies that reported any information about the adverse effects of screening or early diagnosis, but did not attempt to review the adverse effects of hearing aids and cochlear implants. Search results are listed in Appendix 2.

B. Literature Synthesis and Preparation of the Systematic Evidence Review

Twenty-two articles met these inclusion criteria and were abstracted using a standard electronic spreadsheet. From each of the 10 included screening studies (Figure 1, Arrows 1 and 2),[14, 71, 78-85] we abstracted the following information: year of publication, study design (randomized controlled trial, cohort, case-control, controlled case series, or time series comparisons), characteristics of patients studied (risk status, degree and type of hearing loss, age at testing), screening protocol (test used, pass criteria, followup screening and diagnostic

testing), years of data collection, number of patients screened, number with a positive screening test, method used to make the final diagnosis and the age at which it was done, number with confirmed SNHL, age at diagnosis, referrals, compliance with referrals, and age at amplification. We calculated the number of patients with a final diagnosis of bilateral SNHL divided by the number of neonates screened and its inverse, the number needed to screen (NNS) to identify one infant with bilateral SNHL. Where possible, we calculated the NNS in high-risk infants, who would likely be identified by high-risk screening strategies already in place, compared with low-risk infants, who would be identified early only by UNHS. When the information was available we also calculated the NNS for the subset of patients who had complete followup.

For studies of the accuracy of screening tests, we defined *sensitivity* as the number of infants with hearing loss who screened positive divided by the actual number of infants with hearing loss. We defined *specificity* as the number of infants with normal hearing who screened negative divided by the total number of infants with normal hearing. We also calculated the *positive predictive value* as the number of infants with hearing loss who screened positive and later proved to have permanent bilateral SNHL divided by the number of infants who screened positive. The number and type of screening tests administered, and the criteria used to define a positive test, varied among the studies. In most programs, for example, the in-hospital phase of testing had 2 stages (eg, an OAE followed by an ABR, or an ABR repeated once), but other protocols used a single stage (e.g, one OAE or ABR). To be consistent across studies, we defined a screen as positive if, based on whatever tests were done by the time of discharge from the hospital, a referral for repeat testing or audiologic consultation would be recommended.

In most screening methodologies, the gold standard allows for validating the screening tool immediately, in the case of hearing, the accuracy of the gold standard, behavioral and/or audiologic evaluation, depends on the age at which it is performed. Moreover, in the months after discharge from the hospital, audiologic evaluation might be repeated several times before a definitive diagnosis can be made. With input from the Task Force, we defined tests performed in

the hospital during the birth admission as "screening" tests, and defined subsequent testing performed as part of an effort to establish the final diagnosis to be part of the followup evaluation.

It is possible that some cases of hearing loss could develop in the months between birth screening and the gold standard evaluation. As is done for other conditions, for example, Pap smears for cervical cancer and mammography for breast cancer, we classified these as "biological false-negative" results.[86]

For the 8 studies evaluating the effect of screening or early treatment on speech and language outcomes (Figure 1, Arrow 4),[87-93] we abstracted the following information: year of publication, study design, years of data collection, characteristics of patients studied (risk status, degree and type of hearing loss, age at testing, sociodemographic information, family characteristics, cognitive ability), definition of hearing impairment, type of treatment program, and specific tests used to measure receptive and expressive language development, as well as the test scores. Three surveys[85, 94, 95] and one chart review study[96] provided information on adverse effects of early diagnosis and treatment (Figure 1, Arrow 5). We used the USPSTF criteria for grading the quality of studies (Appendix 3) to select the methodologically strongest studies, and to grade the overall evidence for each link in the analytic framework. Study quality is discussed in detail in the Results section.

We constructed a mathematical model of the likely benefits and harms of screening 10,000 newborns. We used the results of the literature review to estimate prevalence, sensitivity and specificity, compliance, treatment effect size, and other parameters of the model.

3. RESULTS

A. Accuracy of Screening Tests

1. Can UNHS Accurately Diagnose Moderate to Severe Sensorineural Hearing Loss?

Since 1995, 10 publications provided new information about the yield of screening and the performance of OAE and ABR in actual screening programs (Table 2). These include a

controlled trial conducted in Wessex, United Kingdom;[79] 5 screening programs at selected hospitals in Rhode Island,[14] Hawaii,[71] Colorado,[81] Texas,[80] and New York;[40] and 4 hospital-based programs, in New Jersey,[82] Whipps Cross, United Kingdom,[97] France,[84] and North Carolina.[85]

2. What are the sensitivity and false negative rate of screening tests?

In most studies, the sensitivity of the screening test could not be assessed, because newborns who passed the initial screen did not undergo further evaluation. Three studies, a controlled trial of UNHS in Wessex,[79] a hospital based program in England, [83] and a report of statewide screening in Rhode Island,[14] provided some information about sensitivity and about the false negative rate of the screening test (1-sensitivity). These studies reported the number of cases missed by screening and eventually diagnosed by other means, but they did not make a comprehensive effort to follow babies who had normal screening test results. The false negative rates were 15%,[79] 6%,[14] and 11%.[83]

A non-randomized, controlled, unblinded trial of UNHS was conducted in the Wessex district of the United Kingdom.[79] Over a period of 3 years, neonatal screening alternated with usual care every 4 to 6 months in 4 maternity hospitals. During the periods of neonatal screening, 21,279 of 25,609 eligible children (83%) had a TEOAE test followed by ABR testing for those with a positive TEOAE test. Newborns with positive ABR results were referred for audiological testing. A total of 28,172 babies were born during periods without neonatal screening. All children in both groups received the existing screening program, the health visitor distraction test (HVDT), a widely used screening test for hearing loss, at about 8 months of age. Children who did not pass the HVDT were also referred for audiological testing. Patients were followed for a period of 18 to 56 months, by which time the final diagnosis, age at diagnosis, and age at initiation of treatment were recorded for all children whose parents had consented to the study.

In the Wessex trial, the sensitivity of TEOAE was higher than that of the HVDT. The two-stage screening protocol identified 23 of 27 (85%) infants who proved to have hearing loss upon followup. In contrast, the HVDT identified less than one-third of newborns who proved to have significant bilateral hearing loss. As the authors note, the results for the HVDT may be *better* than would occur in everyday practice, because the examiners were aware that their performance was being compared to that of the TEOAE. The same applies to the performance of the TEOAE and ABR.[96]

In the Rhode Island study,[14] over 4 years 52,659 infants born at 8 maternity hospitals in the state had a TEOAE and ABR done. The program identified 79 infants with bilateral hearing loss and 27 with unilateral hearing loss. Children with negative test results and no risk factors for hearing loss were assumed to be true negatives. These children did not receive the gold standard evaluation at 6 to 8 months of age. Nevertheless, 5 infants who passed birth screening were diagnosed with bilateral hearing impairment between 5 and 22 months of age (84 bilateral, 27 unilateral SNHL). The sensitivity was therefore 95% for stages 1 and 2. That is, 95% of infants who were confirmed to have hearing impairment failed their newborn hospital screening or the followup outpatient hearing screening at approximately 1 month of age.

3. What are the specificity, false positive rate, and predictive value of screening tests?

In the Wessex trial, the specificity of the combined screening strategy was 98.5%, meaning that 1.5% of normally hearing newborns underwent a full audiologic evaluation because of false positive screening test results.[79] The Rhode Island group reported a specificity of 90% for TEOAE screening in the hospital, meaning that 10% of normal-hearing children underwent the second stage test.[14]

If an infant has a positive result on the screening test, how likely is it that the infant has hearing loss? Because the prevalence of congenital hearing loss is low, there are many more false positives than true positives; as a result, the positive predictive value (PPV) (number of

infants with hearing loss and a positive test divided by the total number testing positive) is also low.

The programs in Table 2 used a two-stage screening protocol, in which an infant who fails the initial test (an OAE or ABR) is retested, either in the hospital or as an outpatient within 12 weeks of discharge, and is referred for audiologic evaluation if he or she fails the second test. The PPV can be calculated for either the first stage or the second stage of screening. If both stages are performed while the infant is in the hospital, the PPV of the second-stage test determines who will be recalled for followup testing as an outpatient. In one good-quality study, the overall PPV for the second-stage screening test was 6.7%.[79] In the well-baby nursery, the PPV was 2.2%, meaning that 1 of every 45 infants referred for outpatient audiologic evaluation eventually proved to have moderate to profound bilateral SNHL. For high risk babies the PPV was 20% (18/90). None of the other studies in Table 2 provided sufficient data to determine the PPV for moderate-profound bilateral PHL. Higher estimates of the PPV reflect the inclusion of unilateral or bilateral mild hearing loss.

In the programs described in Table 2, there was no systematic difference in the performance of TEOAE or ABR when used as the initial test for screening. We identified one good-quality study that measured the sensitivity and specificity of the OAE and ABR using an appropriate behavioral gold standard, visual reinforcement audiometry.[98] In approximately 3,000 high-risk children who underwent neonatal screening and returned for followup testing at 8 to 12 months of age, the two-stage protocol missed 11% of affected ears. The OAE was very sensitive (98%) for severe hearing loss, but was less sensitive (80%) for moderate and profound losses; at this sensitivity, the specificity of the OAE was 80%. For the ABR, sensitivity and specificity were 84% and 90%, respectively. Overall, neonatal testing resulted in a final diagnosis of bilateral moderate to profound SNHL of 1 in 230 high-risk and 1 in 2348 low-risk infants.

These estimates of test performance and yield are probably more reliable than those from actual screening programs in Table 2. In those programs, decisions about diagnosis and treatment are made on the basis of a "diagnostic ABR" performed when the infant is 1 to 6

months of age. The use of this intermediate diagnostic standard facilitates earlier diagnosis and intervention, but may overestimate the number of cases of PHL. In the Wessex trial, the first audiological examination was done when the babies were between 8 and 12 weeks of age. Of 27 diagnosed to have some degree of sensorineural hearing loss, the diagnosis was wrong in 2 (7.4%), and the babies proved to have normal hearing when re-examined at 4 months or 10 months of age.[96] In another study,[83] 5 (29%) of 17 infants initially diagnosed to have moderate PHL were later found to have only mild hearing loss. None of the other studies in Table 2 described follow-up procedures to determine how often the intermediate diagnosis of SNHL was incorrect.

B. Effectiveness of Early Detection

1. Compared to Selective Screening of High-risk Newborns, How Many More Cases are Identified?

No studies have compared the yield of UNHS to the yield of a comparably-staffed, concurrent selective screening program. A logistical problem with selective screening is how to identify low-risk infants in a timely and accurate manner. Low birthweight and admission to a special or intensive care unit are easily identified, but it is not always possible to ascertain a family history of hearing impairment or chromosomal abnormality before hospital discharge.[83] Uncertainty about the proficiency of selective screening makes it difficult to determine how much UNHS adds to the diagnostic yield of screening.

Overall, in the studies in Table 2, screening detected 1 case of moderate to profound SNHL for every 465 to 925 infants screened; from 779 to 2794 low-risk, and 86 to 208 high-risk, newborns were screened to find 1 case. Screening the low-risk or well-nursery population resulted in identification of 5 of the 27 (18.5%) hearing-impaired infants in the Wessex trial,[79] 7 of 22 infants (32%) in the Whipps Cross study,[83] 8 of 15 (53%) hearing-impaired infants in the Hawaii study,[71] and 2 of 6 (33%) in North Carolina.[85] All of the US studies that reported results for low-risk and high-risk groups separately defined “high-risk” as those who had NICU

admission. The New York program examined differences between the NICU and well-baby nursery in detail. Overall, 1 in 884 newborns screened had bilateral hearing loss. In the NICU, where 90% of babies had other risk factors, 1 in 125 had hearing loss. In the well-baby nursery, where 30% had risk factors, 1 in 1042 had hearing loss.[40, 99]

UNHS programs vary in their ability to test all newborns and provide comprehensive follow-up and treatment services:

"Missed" infants. In 5 UNHS programs, from 1% to 4% of newborns were not screened.[14, 71, 78, 80, 82] In the Wessex Trial, 17% of newborns were not successfully screened during their hospitalization.[79] Early discharge or transfer from the hospital, lack of coverage on nights and weekends, and problems with equipment and personnel are the main factors limiting initial coverage.

Universal screening might increase early diagnosis rates in high-risk infants by reducing "miss" rates and increasing compliance with followup testing and treatment. Only one study, in New York, compared miss rates in the NICU to those in the well-baby nursery.[100] In 8 hospitals, miss rates in the well-baby nursery ranged from 1% to 5%, but in the NICUs the miss rates ranged from 3.5% to 26%. Unfortunately, there are no comparison data to determine whether the high miss rates in the NICU are better or worse than those in the New York hospitals that used selective screening. There are also no data on the relative miss rate of selective screening and UNHS for well-baby nursery infants who have a family history or other clinically inapparent risk factors for hearing impairment.

Compliance with followup testing. The programs with the best compliance—Hawaii[71] and Rhode Island[14]—reported that 9% and 13% of infants with an initial positive test failed to return for further testing. Both of these programs had good computerized tracking systems.[14, 101] Other programs reported larger numbers of infants lost to followup. The Texas UNHS program tested 68.5% (1224/1787) of children failing the initial screen; the remaining 31.5% were lost to followup.[80] Mehl reported that 47.8% (1296/2709) of infants in Colorado completed diagnostic evaluation.[81] In the New York program, 72% of inpatient screening failures returned for

followup screening; 76.3% of NICU fails returned versus 71.9% of failures from the well-baby nursery. Only 31% of infants who missed hospital screening returned for outpatient testing. The return rate was 33.5% from the NICU population and 24.3% from the well-baby nursery.[99]

2. How Much Earlier are Children Diagnosed and Treated?

One indicator of the benefit of UNHS is the number of additional cases of significant hearing impairment that are diagnosed early. The rates of early diagnosis and treatment reflect the quality of communication between the screening hospital, consulting audiologists, and service providers. These mechanisms are as important as the reach and accuracy of initial testing to the success of a screening program. Studies should make clear what special mechanisms were put in place to reduce lags between referral and diagnosis, and diagnosis and treatment, that are seen in everyday care.

The Wessex trial did not directly compare the rate of early diagnosis and treatment for UNHS to that of selective screening of high-risk newborns. It did compare UNHS to *no* newborn screening, followed in both groups by HVDT at 8 months of age. In the Wessex trial, for infants with moderate to profound hearing impairment, UNHS increased rates of referral to an audiologist by age 6 months (an increase of 62 per 100,000; 95% CI 19-105/100,000; p=0.006), but did not increase rates of confirmation of diagnosis (p=0.22) or initiation of management within 10 months (p=0.08).[79] Among those with moderate to severe hearing loss, however, screening led to highly significant increases in confirmation and management by 10 months of age. With UNHS, 13 out of 23 (57%) children with moderate or severe impairment were diagnosed by 10 months, whereas during the period of time without UNHS, only 2 out of 14 (14%) with hearing impairment were identified by then. [79] UNHS did not reduce the rate of diagnosis after 18 months, either overall (5/27 for UNHS vs 6/26 for the control group) or in the moderate to severe subgroup, but additional followup may be needed to assess effects on late diagnosis.

How much of the overall benefit in the Wessex trial can be attributed to screening low-risk infants? Of the 27 cases of moderate to profound bilateral hearing loss diagnosed during periods of universal screening, there were 5 who had no risk factors and were in the general (well-baby) nursery. In this low-risk group, 1 in 3597 screened had moderate to profound bilateral SNHL. Overall in the Wessex trial, 59% (16 of 27) babies born during periods of UNHS were diagnosed before 10 months of age. The authors do not break out how many of the *low-risk* cases were diagnosed before that age, but if 59% were, the number needed to screen to detect 1 low-risk infant before 10 months was 7216.[a]

Except for the Wessex trial, the studies in Table 2 were uncontrolled, so the effect on the timing of diagnosis, compared with selective screening of high-risk newborns, cannot be estimated. Some reported decreases in the age at diagnosis as they became more experienced with UNHS. During the 4 years of UNHS in Rhode Island, the mean age of hearing loss detection decreased from 13.3 months at baseline to 5.7 months by year 4.[14] In Hawaii both the average age of hearing-loss identification and fitting with hearing aids decreased as the percent of the population screened by UNHS increased. The mean age of identification was 12 months and the mean age of amplification was 16 months when 19% of the population received screening, compared to an average age of identification of 3 months, and age of amplification of 7 months with almost 95% screening coverage.[101]

Four of the 8 observational studies reported the mean age at the time of treatment. For hearing aid fitting, the mean age for all patients was 5.7 months,[14] 5.8 months,[102] and 7.5 months[78] in the three U.S. studies. In the Whipps Cross study,[83] performed in the United Kingdom, the mean age at amplification was 4.2 months for children who had profound hearing

[a] In the Wessex trial, 11 cases were from the special-care baby unit and 16 from the general ward. Of these 16, 5 had no risk factors for hearing impairment and 11 had at least one risk factor (family history, perinatal infection, anatomical deformity, birth asphyxia, chromosomal abnormality, or exchange transfusion.) Therefore, the overall yield in the general nursery— combining infants with and without risk factors— was one case for every 2,713 screened.

Selective screening would have identified the 11 cases from the special care baby unit (SBCU) and would definitely not have identified the 5 cases from the general nursery who had no risk factors. It is impossible to determine how many of the 11 general-nursery infants with risk factors selective screening would have identified.

loss and 13.8 months for children who had moderate hearing loss. None of these estimates included children who, although screened, did not return for followup testing or treatment (that is, they were not calculated on the appropriate intention-to-treat basis). None of the studies reported information about the technical success of fitting, including how often the hearing aids were used.

The ages of diagnosis in the screening studies were all considerably earlier than those reported in a national survey.[10] The validity of this comparisons is limited, however, because the average ages at detection and amplification exclude cases that might have escaped detection because they were not screened or failed to follow up an abnormal screening test result (ie, they were not calculated on the appropriate intention-to-treat basis). These cases are included in estimates of the time lags in usual care, which are assessed retrospectively and therefore include children diagnosed at a later date, making it likely that the age at diagnosis will appear better in studies of screening than in surveys of usual care.

The evidence that screening leads to earlier treatment could be strengthened in several ways. National surveys of deaf children indicate that the time lag in diagnosis and treatment is shorter for children with profound hearing loss and longest for those with moderate and mild hearing losses.[10] The results reported in studies of screening programs are not directly comparable to the results of surveys because the proportion of patients with profound, severe, moderate, and mild hearing losses are not similar in these groups. The results of the Whipps Cross study,[83] illustrate that this effect is a large one and that results for moderate and profound hearing losses should be reported separately.

One study of universal screening reported the proportion of low-risk infants with moderate to severe impairment who were fitted with a hearing aid or enrolled in language programs by 6 months of age.[99] The median ages at hearing aid fitting were earlier for those in the well baby nursery than NICU, earlier for not-at-risk infants than for at-risk infants, and earlier for those with severe/profound than mild/moderate losses.

3. What are the Potential Adverse Effects of UNHS?

To estimate the *net* benefit of screening—benefits minus harms—it is important to have careful studies of the frequency and seriousness of the adverse effects of screening. For any screening program, false-positive test results may lead to adverse effects. The most serious consequence of a false positive screening test result is a false-positive diagnosis of permanent hearing loss. If treatment is instituted early, a false-positive diagnosis can lead to unnecessary surgery or other treatment in a baby who hears normally.

In the case of newborn hearing screening, the final diagnosis of permanent hearing loss is determined using an extensive audiometric examination by an audiologist using a combination of otolaryngical and audiological examination, diagnostic ABR, and other physiologic testing. This includes behavioral evaluation at 6 to 9 months which confirms electrophysiological diagnosis. In the Wessex trial, the first audiometric examination was done when the babies were between 8 and 12 weeks of age. Of 158 infants who screened positive, 27 were diagnosed to have permanent SNHL; in 2 of these cases (7.4%), however, the diagnosis was wrong, and the babies proved to have normal hearing when re-examined at 4 months or 10 months of age.[96]

Detailed information about the process of establishing the final diagnosis of permanent SNHL is lacking. None of the studies in Table 2 reported how often a definitive diagnosis could be made at the first audiometric examination, and how often it was correct.

Other potentially serious adverse effects of screening include parental misunderstanding and anxiety, and unfavorable labeling. For newborn hearing screening, even a small increased risk of these effects could have a large impact on the net benefit of a screening program. This is because for low-risk infants there are up to 40 false positives for each true case of hearing impairment, and because up to 10 of these false positives do not return for followup, where definitive testing and interviews with health professionals might allay concerns about the baby's

health. No study has attempted to assess directly the effect of a false-positive test result on the parent-infant relationship.

In the Wessex trial, an indirect measure was made in a survey of parents' anxiety and attitudes toward the baby 2 to 12 months after screening.[96] On average, parents whose babies were screened had similar anxiety and attitudes to parents in the unscreened group. It should be noted, however, that, before screening was done, the parents in the screened group received information about the benefits of early identification and gave informed consent for the procedure.

A survey of parents whose children participated in a UNHS program at Whipps Cross hospital in the United Kingdom indicated that 97% of the mothers felt that infant screening was beneficial and 1% reported being very worried about the testing. Among the group of parents whose infants failed the initial screening and received a second test, 3.5% (2/57) of parents reported they were very worried; satisfaction with the testing remained high.[94]

Clemens surveyed 49 mothers of non-NICU infants who failed hospital screening and received a second stage outpatient automated or diagnostic ABR screening test. Two children required additional testing before being diagnosed as hearing normally. Eight percent of mothers (4/49) said they treated their child differently (eg, spoke louder or clapped their hands). Fourteen percent (7/49) reported "lasting anxiety" after the second screening exam. Ninety-four percent (46/49) of the parents of these "false positive" infants approved of UNHS.[85] In a survey at a regional hospital in Logan, Utah (n=169), parents indicated acceptance of newborn screening for their infants, 98.2% of parents said they would give permission for screening, 95.3% would prefer screening even if the baby failed, and 84.9% felt that anxiety caused by failing a screening test would be outweighed by the benefits of early detection.[95]

4. Does Screening and Early Treatment Improve Language and Communication Skills?

No prospective, controlled study directly examined whether newborn hearing screening results in improved speech, language, or educational development. None of the demonstrations of screening described in Table 2 reported the outcomes of treatment for infants identified to have hearing impairment

One cohort study compared language performance in hearing-impaired children who were detected by UNHS to those of an unscreened group (see Evidence Table 1).[87] All subjects in the study were participants in the Colorado Home Intervention Program (CHIP), a multifaceted program that includes fitting for hearing aids and emphasizes home visits and training parents to be involved in helping their deaf child.[103, 104] Lists of participating children were given to the director of the Colorado Newborn Hearing Screening Program, who determined whether the children were born after 1996 in a hospital that employed universal screening and did not have significant cognitive delays. A group of non-screened children born since 1992 (n=25) were selected from hospitals without a universal newborn screening program, after matching on degree of hearing loss (mild, moderate, moderately-severe, profound), cognitive quotient, and age at time of speech and language evaluations. The 2 groups were similar in gender, ethnicity, presence of multiple disabilities, mode of communication, education of primary caregiver, and chronological age.

Language skills were measured using a validated instrument, the Minnesota Child Development Inventory, and measuring vocabulary words used during a videotaped child-parent interaction. When the expressive vocabulary of the screened group and the non-screened group were separately divided into percentiles, some striking findings emerged. The children at the 25th percentile in the screened group had more words (80) than the 75th percentile in the non-screened group (50 words). At the 75th percentile the screened group of children had over 500 words compared to only 50 words for the 75th percentile of non-screened children.

Mean scores for expressive, receptive, and total language were within normal range for screened group and 18 to 21 points higher ($p<.001$) than the unscreened group, whose scores indicated language delay (expressive language 82.9 [SE 3.7] vs. 62.1 [SE 4.3]; receptive

language 81.5 [SE 3.7] vs. 66.8 [SE 4.0]; total language 82.2 [SE 3.3] vs. 64.4 [SE 3.9]). Children identified prior to 6 months (whether in the screened or unscreened group) had a smaller gap between language development and cognitive ability. Language development (total language quotient) was within normal range for 56% of the screened group compared to 24% of the unscreened group. Delayed language development (<70) was present in 24% of the screened group compared to 68% of the unscreened group.

This study has important strengths and weaknesses. The study is important primarily because is the only one that compared the language outcomes of a screened group to that of an unscreened group of patients with congenital hearing loss. The main strengths of the study were that it used relevant, validated measures of language outcomes, controlled for several important potential confounders, and applied the equivalent of an "intention-to-treat" analysis. Using the USPSTF grading system (Appendix 3), however, the study received an overall grade of "poor." The methodologic weaknesses of the study were that assessment of outcomes was not blinded (the examiner could have known whether or not the patient had been detected by screening); the selection of cases was not clearly masked (the director of the screening program may have known the language outcomes when determining who was eligible for the study); and long-term language outcomes and school performance were not measured. Because the screened and unscreened groups were drawn from different hospitals and time periods, it is also possible that the quality of care (not just the timing) provided to the screened group was better than that of the unscreened group.

The study also did not assemble a complete inception cohort of newly-identified patients, a serious flaw that could introduce substantial bias in the results. Ideally, all patients who were diagnosed in a screening hospital and all patients born in nonscreening hospitals would be eligible for inclusion in the study. In this study, eligibility was determined by the availability of an assessment of language outcomes at 2 to 4 years of age, and cases were selected from participants in CHIP rather than from the entire screened and unscreened populations. As noted by the study authors, the database used in this study does not include children who have

congenital hearing loss who did not participate and remain available for followup in CHIP. The proportion of dropouts was not recorded, but could have been measured. Because the unscreened group was diagnosed at later ages, they had less time to benefit from the program or to disenroll than the screened, early-identified group. It is possible that the subgroup of screened patients who remained available for followup at 2 to 4 years of age had a better experience with the program and better outcomes than the full cohort of children diagnosed by screening.

Finally, it is not clear from the study how many of the screened and unscreened subjects are members of the group targeted by UNHS—low-risk infants who would not be diagnosed early by high-risk screening. Moreover, the lack of an inception cohort makes it impossible to determine whether the subjects are representative of this target group.

Studies of preschool children who were identified early or late. Current theory views auditory stimuli during the first 6 months of life as critical to development of speech and language skills.[11-13] This is a critical assumption: if treatment begun at 1 or 2 years of age is as effective as treatment after 6 months, then selective screening of newborns, supplemented by vigilance and appropriate testing if there is parental suspicion, could in theory be as effective as universal screening. (Alternatively, selective screening of newborns could be supplemented by screening of low-risk infants at 7 to 12 months of age.) However, only universal screening can ensure that treatment begins as early as 6 months of age.

Older studies comparing early-identified to late-identified children with impaired hearing consisted of clinical series or case-control studies of highly selected patients, with heterogeneous causes of hearing loss, incompletely defined treatment regimens, and inadequate control for potential confounders.[19] Moreover, these studies are outdated because they classified children diagnosed as late as 3 years of age to be "early-identified."[8] The results of these studies were inconsistent, some supporting and some not supporting the hypothesis that early treatment was associated with better language outcomes. None of these older studies examined the outcome of delayed diagnosis in children who have no risk factors for hearing impairment at birth.

Seven recent cohort studies from 3 programs compared language outcomes for early-identified and late-identified infants (Table 3, Evidence Table 1).[27, 88-93] All of these studies used standardized receptive and expressive tests to evaluate speech and language skills in preschool children, and all reported statistically significant associations between the age at the time of diagnosis and language development at 2 to 5 years of age. Adjusted mean scores for expressive and receptive language were 15 to 20 points higher in groups of children identified and treated early compared to the later identified groups. Although all studies used multivariate analyses, the study populations were composed of convenience samples.

Evidence Table 2 (Quality ratings) summarizes methodologic aspects of these studies. One problem with all of these observational studies is that other confounding factors (that is, factors other than screening and early treatment) may be associated with early diagnosis and with better language outcomes. For example, the early-treated group may have benefited from better family involvement and knowledge or access to better hospitals and providers.

Another problem is that all of these studies report an intermediate outcome—language development in preschool age—rather than the outcomes of primary interest, namely communication, and social, educational, and occupational function later in childhood and in adulthood.

A third problem is that none of them focus on the population of interest—children who would be identified by UNHS but who would escape detection using selective high-risk screening. A fourth characteristic is that none of the studies related specific treatments to outcomes, so it is possible that the differences in language achievement reflect the impact of a single component of the intervention rather than that of a comprehensive treatment plan.

Five studies reported speech and language results for children enrolled in CHIP.[27, 88-90, 92] The most widely cited of these studies compared 72 hearing-impaired children identified prior to 6 months of age to 78 hearing impaired children identified after 6 months.[88] These children were identified from the CHIP database. (It is not clear whether the subjects overlap with those of the Yoshinaga-Itano 2000 study[87] described above.) After adjustment for cognitive function,

children whose hearing losses were identified by 6 months of age demonstrated significantly better receptive, expressive, and total language scores than children identified after 6 months of age. For children with normal cognitive abilities, this language advantage was found across all test ages, communication modes, degree of hearing loss, and socioeconomic strata. Children (with normal cognitive abilities) identified early had language scores at or near their cognitive test scores, whereas children identified after 6 months of age performed, on average, 20 points lower on language scores than cognitive scores. For infants identified and treated by 6 months of age, there was a 20-point-higher mean language quotient in children with normal cognitive quotients compared to infants identified after 6 months (92.2 vs. 71.7[receptive] and 90.5 vs. 68.7 [expressive]). A language quotient of 100 represents functioning at an age-appropriate level; a child with language quotient of 90 is functioning at a level of 90 percent of his/her chronological age. A score of 70.2 means that a 24-month-old communicates at the level of a 17-month-old child, while the language of a 36-month-old resembles that of a 25-month-old. These differences are clinically significant: at 17 to 18 months, a typically developing child has a productive vocabulary of about 50 words and uses few 2-word phrases. At 24 months, a typically developing child has a productive vocabulary of 250 to 300 words (5-6 times the vocabulary he/she had 6 months earlier) and regularly produces sentences of 3 to 5 words. By 36 months, a typically developing child has a vocabulary of over 1000 words. Thus, the child with a language quotient of 70.2 lags 6 months behind peers at 24 months and a year behind at 36 months. The 20-point gap is more than 1 standard deviation lower than normal for age, which would indicate that a child with normal intellect would have the language abilities of a child who had an IQ of 80. Children with low cognitive abilities (cognitive quotient <80), experienced a smaller improvement; the later-identified group had a mean gap between cognitive quotient and language quotient of 10 points. For this group the total language difference was significant ($p=.05$), while receptive and expressive language differences were not significant ($p=.06, .09$).

The main strengths of the study were that it used relevant, validated measures of language outcomes and controlled for several important potential confounders. The study also

has important weaknesses and was rated "poor" using the USPSTF system (Appendix 3). The most important weakness is that the 2 groups—those diagnosed before 6 months and those diagnosed after 6 months—were dissimilar. Patients in the late-identified group were more likely to be cognitively impaired (56% vs. 29%, $p<0.001$), to have severe or worse hearing loss (46% vs. 34%, NS), and to use sign language (54% vs. 46%, NS); their mothers were less likely to have finished high school (52% did not vs. 43%, NS). Language ability was assessed at a later age in the late-identified children than in the early-identified children (56% after 24 months vs. 44% earlier, $p=0.03$). The statistical adjustment used in the analysis did not simultaneously adjust for more than 2 factors and may not have removed the influence of these differences. The data provided in the study were inadequate to determine how much attrition occurred prior to the assessments, and the assessments were not masked.

Additional evidence for the effect of early identification and treatment was provided by a cohort study of 112 children continuously enrolled in a diagnostic early intervention program in Nebraska.[91] It found that after adjustment for family involvement, degree of hearing loss, and nonverbal IQ, children enrolled prior to 11 months had stronger vocabulary and reasoning skills than children enrolled at later ages. The study demonstrated that early identification was related to vocabulary at age 5, but accounted for much less of the variance than family involvement. At age 5, family involvement accounted for 57% of variance in vocabulary, and age of enrollment accounted for 11.5%.

The remaining studies also used multiple regression, but did not adjust for family involvement. For this reason, these studies probably overestimated the association of early enrollment with language development. In one, a retrospective series of 80 children in a home intervention program in Washington State,[93] early enrollment was associated with better language skills at 3 years of age. The relevance to newborn screening is low because only 9 subjects were enrolled before 12 months of age. Mayne and colleagues evaluated factors related to expressive language development in a group of 113 deaf and hard-of-hearing children enrolled in the CHIP program.[92] They reported that expressive vocabulary was higher with

increased age, increased cognitive quotient, identification of hearing loss by the age of 6 months, and having a hearing loss as the only medical condition. The full regression model explained 56% of variance in expressive vocabulary scores.

The studies in Table 3 had several important limitations. The study populations were composed of convenience samples. That is, the studies compared children who were identified early and late by means other than UNHS, rather than children whose age at identification and enrollment was determined primarily by whether or not they were screened. None of the studies had clear criteria for inclusion, none had blinded assessments, and all selected children for inclusion based on the availability of a language assessment between ages 2 to 5. This could introduce bias: early-identified children who remained in the program may have had better results than early-identified children who were not available for followup. Because of these limitations, selection bias cannot be confidently ruled out as an explanation for the findings. Moeller found family involvement an important contributor to language development. Since other studies did not adjust for this factor, they may have overestimated the association of early enrollment with language development. None of the studies provides information on attrition or follow up rates.

5. What are the Potential Adverse Effects of Early Treatment?

The harms of early treatment have not been adequately studied. As noted by the 2nd USPSTF,[19] differing ethical and philosophical attitudes about deaf awareness and culture have led to controversy about the content of early interventions. The argument for early intervention is based on the prevailing theory of language development, which holds that early auditory input is an important precursor of language development. An opposing viewpoint that has been expressed in the literature is that, during infancy, nonverbal communication, joint attention, shared experiences, and mutual understanding are more important precursors of language

development than hearing speech and forming sounds. From this viewpoint, early intervention could be harmful, as expressed in this excerpt from a review article:

> "…there is a risk that if a hearing impairment is detected early and if the parents are recommended to focus on means of communication that the child has the least prerequisites for, the interaction will be altered and gradually impeded. The parents will start to look upon their child as someone with a functional disability, not as a baby with unique competences and possibilities."[105]

Because there are no randomized trials of different management strategies, it is impossible to assess the merits of these concerns. While the studies reviewed above provide no evidence for this hypothesis, they were not adequately designed to detect a harm if one were present. A review of the theoretical literature underlying different attitudes about the treatment of hearing impairment is beyond the scope of this review.

4. DISCUSSION

A. Summary of Benefits and Harms

Table 4 summarizes the benefits and harms of UNHS and selective screening in a hypothetical cohort of 10,000 newborns. We used the results of our literature review to estimate prevalence, sensitivity and specificity, compliance, and the likelihood of being diagnosed and treated before 10 months of age. To estimate PPV, the results of the Wessex trial are used. Although other screening studies found better PPVs, we used this estimate (6.7%) because it comes from the only controlled trial, and is the best evidence about screening.

There are no reliable data by which to estimate how often selective screening misses patients whose risk factors were not detected during hospitalization. We assumed that in a selective screening program, 20% of high-risk infants are never tested in the hospital, versus 10% for

UNHS. There are also no reliable data by which to estimate the probability that a low-risk infant will be diagnosed by 10 months without newborn screening; we estimated this to be 35% in our base case.

With UNHS, an additional 7800 screening tests would be done, resulting in the diagnosis of 6 additional cases of moderate to profound hearing loss diagnosed before 10 months of age. Of these, 3 additional cases would be treated before 10 months of age. Thus, the number needed to screen (NNS) to detect 1 additional case before 10 months would be 1441, and the NNS to treat 1 additional case before 10 months would be 2401. With UNHS, 254 newborns would be referred for audiological evaluation because of false-positive second-stage screening test results, versus 48 for selective screening. Of these, 1 would be falsely diagnosed to have permanent hearing loss at the first post-hospital visit to an audiologist.

Of the 6 additional early-diagnosed, low-risk newborns, how many would actually benefit from early treatment? The data needed to estimate this—the probabilities of a poor language outcome with and without early treatment—are not known. To use a hypothetical example, if 50% of low-risk newborns would have poor language ability if diagnosed after 10 months, and early intervention reduced this by 50%, then the number needed to screen to prevent 1 additional case of delayed language acquisition would be 6771.

B. Conclusions

Table 5 summarizes the evidence for each of the major assumptions underlying the case for UNHS. The quality of the evidence ranged from good to poor; evidence codes are explained in Table 5.

1. There is evidence from population-based surveys that diagnosis and treatment are delayed until after age 1 or 2 in many children who have congenital hearing impairment. This is

particularly true for children who were at low risk for hearing impairment at birth (Level II-2 Evidence).

2. There is anecdotal evidence that these delays result in anxiety and stress on the family or the child (Level III Evidence).

3. Hearing impairment is associated with poor reading and language performance in school and under-employment or unemployment. Delay in diagnosis and treatment is one of many factors associated with these consequences. Later diagnosis is associated with delayed development in vocabulary, grammar, and conversation in preschool children (Level II-3 Evidence). However, the burden of illness has not been directly measured in the main subgroup of children targeted by universal screening—those who have no risk factors for hearing impairment at birth.

4. Technically, OAE and ABR are highly accurate screening tests for congenital hearing impairment (Level II-1 Evidence). In newborns who have risk factors for hearing loss, to achieve a specificity of 90%, the sensitivity of the ABR was 84%. Their accuracy in low-risk newborns has not been carefully studied.

5. The positive predictive value of the two-stage screening protocol (OAE followed by ABR) for the diagnosis of moderate-profound bilateral permanent hearing loss is close to 7% (Level II-2 Evidence). In practice, the protocols used to define a positive test, and consequently the rates of false positive test results, vary widely (up to 10-fold).

6. In expert hands, audiological assessment before 3 months of age results in about one false diagnosis of permanent hearing impairment for every 15 correct diagnoses (Level II-1 Evidence). Studies of screening have not adequately described how often audiological assessment must be repeated, and how often results are indeterminate or false alarms.

7. UNHS increases identification of deaf and hearing-impaired infants compared with selective screening, but several information gaps make it difficult to estimate the difference with

confidence. One controlled study in the United Kingdom suggests that, for infants with moderate and severe hearing losses, screening led to the diagnosis and treatment before 10 months of age of one additional case for every 7216 low-risk newborns screened. (Level II-1 Evidence) Several time series demonstrate that the average age of identification decreased as use of UNHS increased,(Level II-3 Evidence), but these trends also reflect improved coordination of care as well as improved screening coverage.

8. No prospective cohort studies or controlled trials have followed screened and non-screened groups over time to evaluate language outcomes. One cohort study found receptive, expressive, and total language scores of children born in a hospital with UNHS program were 18-21 points higher ($p<.001$) than children born in hospitals without UNHS (Level II-2 Evidence—poor). Several other cohort studies show that, by 2 to 4 years of age, children who have had hearing aids and other therapy in the first 6 months of life had better language skills than those who have had hearing aids and other therapy for shorter periods of time (Level II-2 Evidence—fair to poor). While they are better than older studies, these studies had serious flaws and did not specifically describe the subgroup of children who would be identified by universal screening but not by selective screening.
9. Information on the short and long-term adverse effects of screening and early treatment is inadequate (Level III Evidence). No studies have examined whether early intervention adversely affects the child or the parent-child relationship.

C. Future Research Needs

Gaps in the evidence about the effectiveness of screening are striking. These gaps are best highlighted by comparing UNHS to other screening programs that are considered to be strongly supported by evidence. For some screening proposals, such as mammography or chlamydia testing, randomized trials of screening have proven effects on health outcomes, such as breast cancer mortality or pelvic infections. For other screening programs, such as those for

hyperlipidemia, diabetes, and depression, the efficacy of the proposed treatments has been proven, and the key issue is whether patients who would be identified by screening are sufficiently similar to those included in the randomized trials of treatment. For UNHS, there are no good-quality controlled studies of the clinical benefit of screening, and the research basis for intense early intervention is poor. There are no trials in which early-identified infants are assigned to intense early intervention programs versus expectant management or more conservative approaches.

Gaps in information about UNHS and about alternatives to UNHS remain. While selective screening of high-risk newborns can be justified on the grounds of high prevalence, evidence about the effectiveness of early treatment is no stronger for selective screening than it is for UNHS. Moreover, the ability to identify risk factors in a timely and accurate way is an essential component of the selective screening strategy, but is also poorly studied. Other alternatives to universal screening of newborns include high-risk screening or referral for screening on broad indications at 6 to 12 months of age, when screening tests and audiometric examination are more reliable. These alternatives have not been examined in the literature.

If randomized trials are not done, good-quality longitudinal studies of UNHS versus a high-risk approach or no formal screening are needed to address:

- The gains in case finding by UNHS.
- Quantification of the consequences of false-positive screens and false-negative screens to determine if there are clinically important harms that result from screening.
- Speech, language, and scholastic achievement of deaf and hard-of-hearing children followed over time. These would be examined by method of detection—for example, UNHS, HR, clinical practice and timing (early vs. late) and include information on other important factors that contribute to speech, language and scholastic achievement.
- Comprehensive cost-benefit analyses that include cost of tracking and followup of all children screened. Issues of cost and cost-benefit can be studied from a narrower health care delivery organization's point of view and/or a broader societal perspective. In integrated

delivery systems such as HMOs, issues of total and marginal direct medical costs and patterns of care could be addressed.

- Registries that are part of integrated service delivery systems should be encouraged at the city, county, or state levels. These registries could play a vital role in routinely collecting information on known and potential confounders as well as speech and language outcome data. In order to answer the service delivery and developmental questions an integrated screening, an evaluation and followup system with universal access is needed. The requirements are similar to those for congenital metabolic defects such as phenylketonuria (PKU) or hypothyroidism.
- Large-scale screening, intervention, and longitudinal followup studies with developmental outcome measurement for UNHS. Early Hearing Detection and Intervention and Health Resources and Services Administration programs might provide funding and vehicles for this research.

REFERENCES

1. Ruben RJ. Effectiveness and efficacy of early detection of hearing impairment in children. Acta Otolaryngol 1991;482(Suppl.):127-31.
2. Rach GH, Zielhuis GA, van den Broek P. The influence of chronic persistent otitis media with effusion on language development of 2- to 4- year-olds. Int J Pediatr Otorhinolaryngol 1988;15:253-61.
3. Zinkus PW, Gottlieb MI. Patterns of Perceptual and Academic Deficits Related to Early Chronic Otitis Media. Pediatrics 1980;66(2):246-53.
4. Moeller MP, Osberger MJ, Eccarius M. Receptive language skills. Language and Learning Skills of Hearing-Impaired Children; 1986. p. 41-53.
5. Carney AE, Moeller MP. Treatment efficacy: hearing loss in children. J Speech Lang Hear Res 1998;41(1):S61-84.
6. National Institutes of Health. Early identification of hearing impairment in infants and younger children. Rockville, Maryland; 1993.
7. Fonseca S, Forsyth H, Grigor J, Lowe J, MacKinnon M, Price E, et al. Identification of permanent hearing loss in children: are the targets for outcome measures attainable?. Br J Audiol 1999;33(3):135-43.
8. Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S. A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment. Health Technol Assess 1997;1(10):i-iv, 1-176.
9. Meadow-Orlans KP, Mertens DM, Sass-Lehrer MA, Scott-Olson K. Support services for parents and their children who are deaf or hard of hearing: a national survey. Am Ann Deaf 1997;142(4):278-93.
10. Harrison M, Roush J. Age of suspicion, identification, and intervention for infants and young children with hearing loss: a national study. Ear Hear 1996;17(1):55-62.
11. Kuhl PK, Williams KA, Lacerda F, Stephens KN, Lindbloom B. Linguistic experience alters phonetics perception in infants by six months of age. Science 1992;255:606-08.

12. Gopnik A, A. M, Kuhl PK. The scientist in the crib: Minds, brains, and how children learn. New York: William Morrow and Company; 1999.
13. Sininger YS, Doyle KJ, Moore JK. The case for early identification of hearing loss in children. Auditory system development, experimental auditory deprivation, and development of speech perception and hearing. Pediatr Clin North Am 1999;46(1):1-14.
14. Vohr BR, Carty LM, Moore PE, Letourneau K. The Rhode Island Hearing Assessment Program: experience with statewide hearing screening (1993-1996). J Pediatr 1998;133(3):353-7.
15. Finitzo T, Diefendorf AO. The state of the information... evidence gathering in infant hearing programs. Am J Audiol 1997;6(3):91-4.
16. White KR, Vohr BR, Behrens TR. Universal newborn hearing screening using transient evoked otoacoustic emissions: results of the Rhode Island haring Assessment Project. Semin Hear 1993;14(1):18-29.
17. Joint Committee on Infant Hearing. Paper presented at: Joint committee on infant hearing year 2000 position statement: principles and guidelines for early hearing detection and intervention programs, February 14, 2000.
18. Fortnum H, Davis A. Epidemiology of permanent childhood hearing impairment in Trent Region, 1985-1993. Br J Audiol 1997;31(6):409-46.
19. U.S. Preventive Services Task Force. Screening for hearing impairment. In: U.S. Preventive Services Task Force. Guide to clinical preventive services,. 2nd edition. ed. Baltimore: Williams & Wilkins; 1996. p. p 393-405.
20. Grandori F, Lutman ME. European Consensus Statement on Neonatal Hearing Screening. Finalised at the European Consensus Development Conference on Neonatal Hearing Screening, 15-16 May 1998, Milan. Int J Pediatr Otorhinolaryngol 1998;44:309-10.
21. Stein LK. Factors influencing the efficacy of universal newborn hearing screening. Pediatr Clin North Am 1999;46(1):95-105.

22. Maki-Torkko EM, Lindholm PK, Vayrynen MRH, Leisti JT, Sorri MJ. Epidemiology of moderate to profound childhood hearing impairments in northern Finland. Any changes in ten years? Scand Audiol 1998;27(2):95-103.
23. Karikoski JO, Marttila TI. Prevalence of childhood hearing impairment in southern Finland. Scand Audiol 1995;24(4):237-41.
24. Parving A, Jensen JH. Prevalence of permanent childhood hearing impairment -- its role in audit of local paediatric hearing health services. J Audiol Med 1998;7(2):100-8.
25. Boyle CA, Yeargin-Allsopp M, Doernberg NS, Holmgreen MS, Murphy CC, Schendel DE. Prevalence of selected developmental disabilities in children 3-10 years of age: the metropolitan Atlanta developmental disabilities surveillance program, 1991. MMWR, CDC Surveillance Summaries 1996;45(SS-2):1-14.
26. Van Naarden K, Decoufle P, Caldwell K. Prevalence and Characteristics of Children with Serious Hearing Impairment in Metropolitan Atlanta, 1991-1993. Pediatrics; 1999. p. 570-75.
27. Yoshinaga-Itano C, Apuzzo ML. The development of deaf and hard of hearing children identified early through the high-risk registry. Am Ann Deaf 1998;143(5):416-24.
28. Karchmer MA, Allen TE. The functional assessment of deaf and hard of hearing students. Am Ann Deaf 1999;144(2):68-.
29. Holt JA. Stanford Achievement Test-8th edition;: reading comprehension subgroup results. Am Ann Deaf Ref Iss 1993;138:172-75.
30. Allen TE. Patterns of academic achievement among hearing impaired students: 1974 and 1983. Deaf Children in America 1986:161-206.
31. Quigley SO. Environment and communication in the language development of deaf children. In: Bradford LJ, Hardy WG, editors. Hearing and hearing impairment. New York, NY: Grune and Stratton; 1979.
32. Wrightstone JW, Aronow MS, Moskowitz S. Developing reading test norms for deaf children. Am Ann Deaf 1963;108:311-16.

33. Furth HG. A comparison of reading test norms of deaf and hearing children. Am Ann Deaf 1966;111:461-2.
34. Gravel JS, Wallace IF, Ruben RJ. Auditory consequences of early mild hearing loss associated with otitis media. Acta Otolaryngol 1996;116(2):219-21.
35. Bess FH, Dodd-Murphy J, Parker RA. Children with minimal sensorineural hearing loss: prevalence, educational performance, and functional status. Ear Hear 1998;19(5):339-.
36. Tharpe AM, Bess FH. Minimal, progressive, and fluctuating haring losses in children: Characteristics, identification, and managment. Pediatr Clin North Am 1999;46(1):65-78.
37. Dirckx JJ, Daemers K, Somers T, Offeciers FE, Govaerts PJ. Numerical assessment of TOAE screening results: currently used criteria and their effect on TOAE prevalence figures. Acta Otolaryngol 1996 Sep;116(5):672-9.
38. Fowler KB, McCollister FP, Dahle AJ, Boppana S, Britt WJ, Pass RF. Progressive and fluctuating sensorineural hearing loss in children with asymptomatic congenital cytomegalovirus infection. J Pediatr 1997;130(4):624-30.
39. Lasky RE, Wiorek L, Becker TR. Hearing loss in survivors of neonatal extracorporeal membrane oxygenation (ECMO) therapy and high-frequency oscillatory (HFO) therapy. J Am Acad Audiol 1998;9(1):47-58.
40. Prieve BA, Stevens F. The New York State universal newborn hearing screening demonstration project: introduction and overview. Ear Hear 2000;21(2):85-91.
41. Widen JE, Folsom RC, Cone-Wesson B, Carty L, Dunnell JJ, Koebsell K, et al. Identification of neonatal hearing impairment: hearing status at 8 to 12 months corrected age using a visual reinforcement audiometry protocol. Ear Hear 2000;21(5):471-87.
42. Moore JM, Thompson G, Folsom RC. Auditory responsiveness of premature infants utilizing visual reinforcement audiometry (VRA). Ear Hear 1992;13:187-94.
43. Folsom RC, Diefendorf AO. Physiologic and behavioral approaches to pediatric hearing assessment. Pediatr Clin North Am 1999;46(1):107-20.

44. Stredler-Brown A, Arehart KH. Universal newborn hearing screening: impact on early intervention services. Volta Review In Press.
45. Grandori F, Lutman M. The European Consensus Development Conference on Neonatal Hearing Screening (Milan, May 15-16, 1998). Am J Audiol 1999;8:19-20.
46. Brookhouser PE. Sensorineural hearing loss in children. [Review]. Pediatr Clin North Am 1996;43(6):1195-216.
47. Eilers RE, Berlin C. Advances in early detection of hearing loss in infants. Curr Probl Pediatr 1995;25(2):60-6.
48. Finitzo T, Crumley WG. The role of the pediatrician in hearing loss. From detection to connection. [Review]. Pediatr Clin North Am 1999;46(1):15-34, ix-x.
49. Martin WH, Schwegler JW, Gleeson AL, Shi YB. New techniques of hearing assessment. [Review]. Otolaryngol Clin North Am 1994;27(3):487-510.
50. Oudesluys-Murphy AM, van Straaten HL, Bholasingh R, van Zanten GA. Neonatal hearing screening. [Review]. Eur J Pediatr 1996;155(6):429-35.
51. Silverman CA. Audiologic assessment and amplification. Prim Care 1998;25(3):545-81.
52. White KR, Maxon AB. Early identification of hearing loss: implementing universal newborn hearing screening programs. Logan, UT: National Center for Hearing Assessment and Management (NCHAM).
53. Erenberg A, Lemons J, Sia C, Trunkel D, Ziring P. Newborn and infant hearing loss: detection and intervention.American Academy of Pediatrics. Task Force on Newborn and Infant Hearing, 1998- 1999. Pediatrics 1999;103(2):527-30.
54. Robinshaw HM. Acquisition of speech, pre- and post-cochlear implantation: longitudinal studies of a congenitally deaf infant. Eur J Disord Commun 1996;31(2):121-39.
55. Robinshaw HM. The pattern of development from non-communicative behaviour to language by hearing impaired and hearing infants. Br J Audiol 1996;30(3):177-98.
56. Robinshaw HM. Early intervention for hearing impairment: differences in the timing of communicative and linguistic development. Br J Audiol 1995;29(6):315-34.

57. White SJ, White REC. The effects of hearing status of the family and age of intervention on receptive and expressive oral language skills in hearing-impaired infants. Development of Language and Communication Skills in Hearing-impaired Children. p. 9-24.
58. Watkins S. Long term effects of home intervention with hearing-impaired children. Am Ann Deaf 1987;132:267-71.
59. Musselman CR, Wilson AK, Lindsay PH. Effects of early intervention on hearing impaired children. Exceptional Children 1988;55(3):222-8.
60. Skinner MW. The hearing of speech during language acquisition. Otolaryngol Clin North Am 1978;11(3).
61. McFarland W, H., Simmons FB. The importance of early intervention with severe childhood deafness. Pediatr Ann 1980;9(1):13-19.
62. Yoshinaga-Itano C, Sedey AL. Speech development of deaf and hared-of-hearing children in early childhood: inter-relationships with language and hearing. Volta Review In Press.
63. Abrams MJ, Pensak ML, Buhrer K. Perspectives on a state enacted hearing screening assessment program in the newborn population. Am J Otol 1997 May;18(3):368-72.
64. Apostolopoulos NK, Psarommatis IM, Tsakanikos MD, Dellagrammatikas HD, Douniadakis DE. Otoacoustic emission-based hearing screening of a Greek NICU population. Int J Pediatr Otorhinolaryngol 1999;47(1):41-8.
65. Chen SJ, Yang EY, Kwan ML, Chang P, Shiao AS, Lien CF. Infant hearing screening with an automated auditory brainstem response screener and the auditory brainstem response. Acta Paediatr 1996;85(1):14-8.
66. Comerford DG, Watson C, Khan MS, Hussain SS. The Bradford and Airedale baby hearing project. An assessment of the impact of screening on the earlier detection of infant hearing loss. Clinical Otolaryngology & Allied Sciences 1995;20(6):536-9.
67. Elmy HAE, Jurkovicova J, Aghova L. Methodical possibilities of early hearing screening in prevention of hearing loss. Studia Psychologica 1994;36(5):369-73.

68. Hunter MF, Kimm L, Cafarlli Dees D, Kennedy CR, Thornton AR. Feasibility of otoacoustic emission detection followed by ABR as a universal neonatal screening test for hearing impairment. Br J Audiol 1994;28(1):47-51.
69. Daemers K, Drickx JD, Van Driessche K, Somers T, Offeciers FE, Govaerts PJ. Neonatal hearing screening with otoacoustic emissions: an evaluation. Acta Otorhinolaryngol Belg 1996;50(3):203-9.
70. Lutman ME, Davis AC, Fortnum HM, Wood S. Field sensitivity of targeted neonatal hearing screening by transient-evoked otoacoustic emissions. Ear Hear 1997;18(4):265-76.
71. Mason JA, Herrmann KR. Universal infant hearing screening by automated auditory brainstem response measurement. Pediatrics 1998;101:221-28.
72. Oudesluys-Murphy AM, Harlaar J. Neonatal hearing screening with an automated auditory brainstem response screener in the infant's home. Acta Paediatr 1997;86(6):651-5.
73. Sorenson P. Universal hearing screening in the NICU: the Loma Linda University Children's Hospital experience. Neonatal Network - Journal of Neonatal Nursing 1998;17(7):43-8.
74. Stewart DL, Pearlman A. Newborn hearing screening. J Ky Med Assoc 1994;92(11):444-9.
75. Sutton GJ, Scanlon PE. Health visitor screening versus vigilance: outcomes of programmes for detecting permanent childhood hearing loss in West Berkshire. Br J Audiol 1999;33(3):145-56.
76. Meyer C, Witte J, Hildmann A, Hennecke KH, Schunck KU, Maul K, et al. Neonatal screening for hearing disorders in infants at risk: incidence, risk factors, and follow-up. Pediatrics 1999;104(4 Pt 1):900-4.
77. Mahoney, Eichwald. 1987.
78. Prieve B, Dalzell L, Berg A, Bradley M, Cacace A. The New York State universal newborn hearing screening demonstration project: outpatient outcome measures. Ear Hear 2000;21(2).

79. Wessex Universal Neonatal Hearing Screening Trial Group. Controlled trial of universal neonatal screening for early identification of permanent childhood hearing impairment. Wessex Universal Neonatal Hearing Screening Trial Group. Lancet 1998;352:1957-64.
80. Finitzo T, Albright K, O'Neal J. The Newborn with Hearing Loss: Dectection in the Nursery. Pediatrics 1998;102(6):1452-60.
81. Mehl A, Thomson V. Newborn hearing screening: the great omission. Pediatrics 1998;E4:101.
82. Barsky-Firkser L, Sun S. Universal newborn hearing screenings: a three-year experience. Pediatrics 1997;99(6):E4.
83. Watkin PM. Outcomes of neonatal screening for hearing loss by otoacoustic emission. Arch Dis Child Fetal Neonatal Ed. 1996;75(3):F158-68.
84. Aidan D, Avan P, Bonfils P. Auditory screening in neonates by means of transient evoked otoacoustic emissions: a report of 2,842 recordings. Ann Otol Rhinol Laryngol 1999 Jun;108(6):525-31.
85. Clemens CJ, Davis SA, Bailey AR. The "false positive" in universal newborn hearing screening. Pediatrics 2000.
86. Eddy DM. Common Screening Tests. Philadelphia: American College of Physicians; 1991.
87. Yoshinaga-Itano C, Coulter D, Thomson V. The Colorado Newborn Hearing Screening Project: effects on speech and language development for children with hearing loss. J Perinatol 2000;20(8 Pt 2):S132-7.
88. Yoshinaga-Itano C, Sedey AL, Coulter DK, Mehl AL. Language of early- and later-identified children with hearing loss. Pediatrics 1998;102(5):1161-71.
89. Yoshinaga-Itano C, Apuzzo ML. Identification of hearing loss after age 18 months is not early enough. Am Ann Deaf 1998;143(5):380-7.
90. Apuzzo M-rL, Yoshinaga-Itano C. Early identification of infants with significant hearing loss and the Minnesota child development inventory. Semin Hear 1995;16(2):124-39.

91. Moeller MP. Early intervention and language development in children who are deaf and hard of hearing. Pediatrics 2000.
92. Mayne AM, Yoshinaga-Itano C, Sedey AL, Carey A. Expressive vocabulary development of infants and toddlers who are deaf or hard of hearing. Volta Review 2000.
93. Calderon R, Naidu S. Further support of the benefits of early identification and intervention with children with hearing loss. Volta Review 2000.
94. Watkin P, Baldwin M, Dixon, R, Beckman A. Maternal anxiety and attitudes to universal neonatal hearing screening. Br J Audiol 1998:27-37.
95. Barringer DG, Mauk GW. Survey of parents perceptions regarding hospital-based newborn hearing screening. Audiol Today 1997;9:18-19.
96. Kennedy CR. Controlled trial of universal neonatal screening for early identification of permanent childhood hearing impairment: coverage, positive predictive value, effect on mothers and incremental yield. Wessex Universal Neonatal Screening Trial Group. Acta Paediatr Suppl 1999;88(432):73-5.
97. Watkin PM. Neonatal otoacoustic emission screening and the identification of deafness. Arch Dis Child Fetal Neonatal Ed. 1996;74(1):F16-25.
98. Norton SJ, Gorga MP, Widen JE, Folsom RC, Sininger Y, Cone-Wesson B, et al. Identification of neonatal hearing impairment: evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem response test performance. Ear Hear 2000;21(5):508-28.
99. Dalzell L, Orlando M, MacDonald M, A. B, Bradley M, Cacace A, et al. The New York State universal newborn hearing screening demonstration project: ages of hearing loss identification, hearing aid fitting, and enrollment in early intervention. Ear Hear 2000;21(2):118-30.
100. Spivak L, Dalzell L, Berg A, Bradley M, Cacace A. The New York State universal newborn hearing screening demonstration project: inpatient outcome measures. Ear Hear 2000;21(2):92-.

101. Johnson JL, Kuntz NL, Sia CC, White KR, Johnson RL. Newborn hearing screening in Hawaii. Hawaii Med J 1997;56(12):352-5.

102. Mason S, Davis A, Wood S, Farnsworth A. Field sensitivity of targeted neonatal hearing screening using the Nottingham ABR Screener. Ear Hear 1998;19(2):91-102.

103. Thomson V. The Colorado Newborn Hearing Screening Project. Am J Audiol 1997;6(3):74-7.

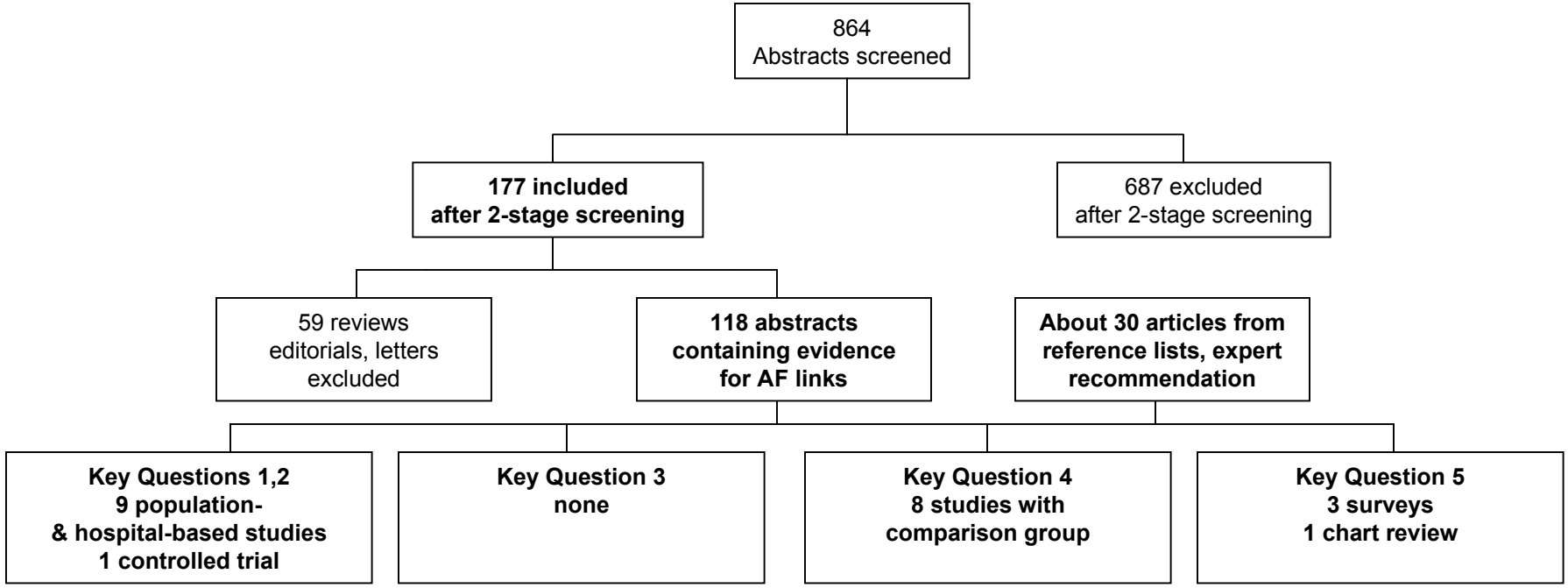
104. Thomson V, Rose LB, O'Neal J, Finitzo T. Statewide implementation of universal newborn hearing screening. Semin Hear 1998;19(3):287-300.

105. Priesler G. The development of communication and language in deaf and severely hard of hearing children: implications for the future. Int J Pediatr Otorhinolaryngol 1999;49 Suppl 1:S39-43.

Appendix 1. Search Strategy

Set	Search
1	exp hearing disorders/
2	infant/or infant, newborn/
3	1 and 2
4	limit 3 to human/
5	limit 4 to English language/
6	4 not 5
7	limit 6 to abstracts
8	5 or 7
9	exp mass screening/
10	screen$.tw.
11	exp hearing tests/
12	9 or 10 or 11
13	8 and 12
14	cochlear implants/
15	exp hearing aids/
16	exp manual communication/
17	exp rehabilitation of hearing impaired/
18	esp hearing disorders/dt,rh,su,th
19	14 or 15 or 16 or 17 or 18
20	8 and 12
21	13 or 20
22	exp hearing disorders/
23	limit 22 to human
24	limit 23 to english language
25	limit 24 to (preschool child <2 to 5 years> or child <6 to 12 years> or adolescence <13 to 18 years>)
26	19 and 25
27	exp evaluation studies/
28	follow-up studies/
29	meta analysis/
30	exp clinical trials/
31	27 or 28 or 29 or 30
32	26 and 31
33	limit 26 to (controlled clinical trial or guideline or meta analysis or mulitcenter study or practice guideline or randomized controlled trial or review, multicase)
34	32 or 33
35	21 or 34

Appendix 2. Literature search results



Appendix 3. US Preventive Services Task Force Quality Rating

Randomized Controlled Trials and Cohort Studies

Criteria:

- Initial assembly of comparable groups.
  a. For RCTs: adequate randomization, including first concealment and whether potential confounders were distributed equally among groups.
  b. For cohort studies: consideration of potential confounders with either restriction or measurement for adjustment in the analysis; consideration of inception cohorts.
- Maintenance of comparable groups (includes attrition, cross-overs, adherence, contamination).
- Levels of follow-up: differential loss between groups;overall loss to follow-up
- Measurements: equal, reliable, and valid, and including masking of outcome assessment.
- Clear definition of interventions.
- Important outcomes considered.
- Analysis:
  a. For RCTs: intention-to-treat analysis
  b. For cohort studies: adjustment for potential confounders.

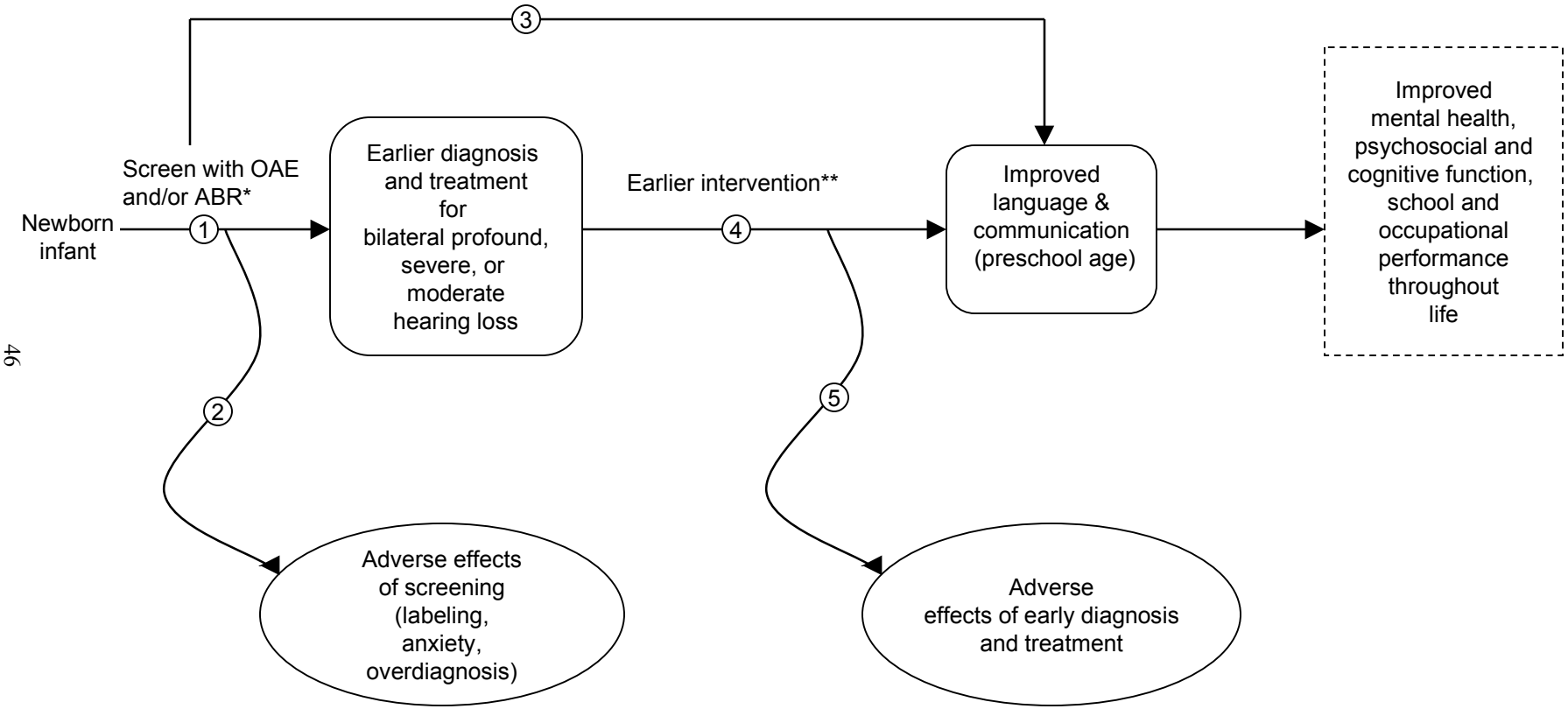
Definition of ratings based on above criteria:

Good: Meets all criteria: comparable groups are assembled initially and maintained throughout the study; follow-up at least 80 percent; reliable and valid measurement instruments applied equally to the groups; interventions clearly defined; important outcomes are considered; and appropriate attention to confounders in analysis. In addition, for RCTs, intention-to-treat analysis is used.

Fair: Generally comparable groups assembled initially but some question remains whether some (although not major) differences occurred in follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention-to-treat analysis is done for RCTS.

Poor: Groups assembled initially are not close to being comparable or maintained throughout the study; measurement instruments are unreliable or invalid or not applied at all equally among groups; outcome assessment not masked; and key confounders are given little or no attention. For RCTs, no intention-to-treat analysis.

Figure 1. Analytic Framework



* OAE = Otoacoustic emissions, ABR= auditory brainstem response.

** hearing aids or other amplification, early American Sign Language and/or English instruction, speech & language therapy, family education & support.

Figure 2. Key questions in analytic framework.

1. Can UNHS accurately diagnose moderate to severe sensorineural hearing impairment? (Arrow 1)
    - What are the sensitivity and false negative rate of screening tests?
    - What are the specificity, false positive rate, and predictive value of screening tests?
    - Compared to selective screening of high-risk newborns, how many more cases are identified?
    - How much earlier are children diagnosed and treated?
2. What are the potential adverse effects of universal screening? (Arrow 2)
3. Does screening lead to improved language and communication skills? (Arrow 3)
4. Does treatment prior to 6 months lead to improved language and communication in infants who would not be diagnosed that early in a selective, high-risk screening program? (Arrow 4)
5. What are the potential adverse effects of early treatment? (Arrow 5)

Table 1. Risk factors for sensorineural hearing loss in newborns.*

1. NICU admission for 2 or more days

2. Usher's Syndrome, Waardenburg's Syndrome, or findings associated with other syndromes known to include hearing loss

3. Family history of hereditary childhood sensorineural hearing loss

4. Congenital infections such as toxoplasmosis, bacterial meningitis, syphilis, rubella, cytomegalovirus, and herpes

5. Craniofacial anomalies, including morphologic abnormalities of the pinna and ear canal

*Joint Committee on Infant Hearing criteria for identifying infants at high risk for hearing loss.[17]

Table 2. Studies of universal newborn hearing screening

Author Year	Description (Quality Rating)	Screening Tests	# Screened/ # Available (% Screened)	Yield & NNS to find 1 case of bilateral PHL	# Positive Screen (%) # *Follow-up* *%Lost to Follow-up*	Definition of High-Risk	# Low-Risk Identified/ # Screened NNS	#High Risk Identified/ #Screened NNS	True positive /Total positive tests PPV	PPV (LR) Calculated by Overall NP rate	PPV (HR) Calculated by Overall NP rate
Wessex 1998[79]	Controlled, nonrandomized trial from 4 hospitals of 25,609 from 10/93 to 10/96 (Good)	TEOAE followed by ABR	21,279/25,609 (83%)	23/21,279 925	342 (1.6%) *NR* *NR*	NIH Criteria	7/19,555 2,794	20/1,724 86	23/342 6.7%	2.2%*	72.5%*
Prieve 2000[40]	State-wide demonstration project from 7 perinatal centers, 8 hospitals in New York (Good)	TEOAE followed by TEOAE or ABR in birth admission; TEOAE, ABR at 4-6 weeks (stage 2)	69,766/71,922 (97%)	49/69,736 1422	4699 (6.5%) 1st stage See footnote* *43.4%*	NICU infants	33/NR 2041†	52/NR 208†	NR 4.5% (stage 1) 22.1% (stage 2)	2.2%	12.5%
Vohr 1998[14]	Cohort from 8 maternity hospitals in Rhode Island from 1/93 to 12/96 (Fair)	TEOAE followed by: ABR (HR infants) TEOAE and ABR in 2-6 weeks (LR infants)	52,659/53,121 (99%)	79/52,659 666†	5,397 (10.2%) 1st stage 677 (1.3%) 2nd stage *4,575* *15.2%*	NICU infants	61/47,529 779 ‡	50/5,130 103 ‡	79/5,397 1.5% *(stage 1)* 79/677 11.7% *(stage 2)*	9.9%*	75%*
Finitzo 1998[80]	Cohort from 9 Texas hospitals from 1/94 to 6/97 (Fair)	ABR or TEOAE in birth admission followed by either ABR or TEOAE at 1-8 weeks	52,508/54,228 (97%)	20/17,105 855† §	1,787 (3.4%) *1,224* *31.5%*	NR	NR	NR	113/1,787 6.3% or 113/1,224 9.2% (stage 1) Total positive NR (stage 2)	NR	NR
Barsky-Firsker 1997[82]	Hospital-based series at Saint Barnabas Medical Center, New Jersey 1/93 to 12/95 (Fair)	ABR by audiologists. (One-stage)	15,749/16,229 (97%)	*NR*	485 (3.1%) NR NR	NICU infants	29/14,014 483	23/1,735 75	52/485 10.7%	6.7%*	42.8%*
Watkin 1996[83]	Hospital-based series at Whipps Cross Hospital, England (Fair)	TEOAE followed by TEOAE and ABR within 4 weeks	11,606/14,353 (81%)	19/11,606 755	337 (2.9%) *290* *14%*	Risk Factors and/or NICU	7/NR	13/NR	18/337 5.3% or 18/290 6.2% (stage 1)		
Mehl 1998[81]	Cohort from 26 hospitals in Colorado from 1992 to 1996 (Poor)	19 ABR, 1 TEOAE, 6 ABR Follow-up screen not reported	41,796 /NR	NR	2709 (6.5%) *1,296* *52.2%*	NR	NR	NR	75/2,709 2.8% or 75/1,296 5.8% (stage 1)	NR	NR

ABR indicates automated auditory brainstem response; TEOAE, transient evoked otoacoustic emissions; NIH, National Institutes of Health Consensus Development Conference; LR, low-risk; HR, high-risk; NICU, neonatal intensive care unit; FH, family history; NR, not reported; NNS, number needed to screen; PHL, permanent hearing loss (moderate or worse)

*Reported different rates for misses and fails

†Includes mild, bilateral hearing loss

‡Includes unilateral hearing loss

§Data reported for 1996 only

Table 2. Studies of universal newborn hearing screening

Author Year	Description (Quality Rating)	Screening Tests	# Screened/ # Available (% Screened)	Yield & NNS to find 1 case of bilateral PHL	# Positive Screen (%) *# Follow-up %Lost to Follow-up*	Definition of High-Risk	# Low-Risk Identified/ # Screened NNS	#High Risk Identified/ #Screened NNS	True positive /Total positive tests PPV	PPV (LR) Calculated by Overall NP rate	PPV (HR) Calculated by Overall NP rate
Aidan, 1999[84]	Hospital-based series in Paris, France of infants in normal newborn nursery (Poor)	TEOAE in 48 hours; TEOAE within 4 weeks	1,421/1,727 (82%)	2/1,421 711	238 (16.7%) *123* *48.3%*	hypoxemia, hyperbili-rubinemia, FH	2/1,421 711	NR	2/238 0.8% or 2/123 1.6% (stage 1) 2/9 22.2% (stage 2)	2/238 (stage 1) 0.8% 2/9 (stage 2) 22.2%	NR
Clemens 2000[85]	Hospital-based series in North Carolina (Women's Hospital of Greensboro) from 7/98 to 6/99 (Poor)	ABR followed by re-test for fails (stage 1a) or ABR (stage 1b); outpatient ABR and diagnostic ABR (stage 2)	5010/5034 (99.5%)	*NR*	103/5,054 *85* *17.5%*	NICU infants	NR	4/454 114	9/103 (stage1a,b) 8.7% 9/15 (stage 2) 60% unilateral and bilateral losses	9/103 8.7%	4/5 80% (one patient lost to follow-up)
Mason 1998[102]	Series of infants born at Kaiser, Honolulu from 3/92 to 2/97 (Poor)	ABR	10,372/10,773 (96%)	12/10,372 864	415 (4.0%) *362* *12.8%*	NICU infants	5/8,971 1,794	7/1,401 200	15/415 3.6% or 15/362 4.1% (stage 1) total positive NR *(stage 2)*	2.2%*	12.5%*

ABR indicates automated auditory brainstem response; TEOAE, transient evoked otoacoustic emissions; NIH, National Institutes of Health Consensus Development Conference; LR, low-risk; HR, high-risk; NICU, neonatal intensive care unit; FH, family history; NR, not reported; NNS, number needed to screen; PHL, permanent hearing loss (moderate or worse)

*Reported different rates for misses and fails

†Includes mild, bilateral hearing loss

‡Includes unilateral hearing loss

§Data reported for 1996 only

Table 3. Cohort studies reporting language outcomes

Study, Year	Selection of Subjects	Comparability and Maintenance of Early vs. Late Groups	Adjustment for Confounders	Results
Studies from Colorado Home Intervention Program (CHIP)				
Apuzzo, 1995[90]	Convenience sample of 69 high-risk infants diagnosed between 2 and 25 months of age. Children with severe cognitive delay were excluded.	Late-identified group was more likely to have severe-profound hearing loss (65% vs. 50%). No report of attrition or followup rates.	One-way ANOVA did not adjust for SES, family involvement, or other potential confounders.	At 40 months of age, infants identified before 2 months of age had higher mean Minnesota Child Development Inventory (MCDI) scores for expressive language ($p<0.01$)
Yoshinaga-Itano, 1998[89]	Convenience sample of 40 high-risk infants, divided into those identified and treated before 6 months of age (n=15) and those treated after 18 months (n=25). Children with severe cognitive delay were excluded (DQ<60).	Late-identified group was more likely to have severe to profound hearing loss, (52% vs 47%). No report of attrition or followup rates.	Gender, severity of hearing loss, cognitive function, and other disabilities were examined in 2-way ANCOVAs, not in a multiple regression (no simultaneous adjustment for multiple confounders).	At 40 months, infants identified before 6 months of age had better adjusted mean MCDI scores for expressive language (81.1 vs 64.3, $p<0.05$) and receptive language (84.4 vs. 70.1, $p<0.05$)
Yoshinaga-Itano, 1998[27]	Convenience sample of 82 infants, 19 to 36 months of age with mild to profound SNHL, divided into those identified before 6 months of age (n=34) and between 7-18 months of age (n=48). Early group identified by HRR; late group by usual care. Children with severe cognitive delay were excluded (DQ<60).	Late identified group was more likely to have severe to profound hearing loss, (77% vs 42%). No report of attrition or followup rates.	Gender, severity of hearing loss, cognitive function, and other disabilities were examined in 2-way ANCOVAs, not in a multiple regression (no simultaneous adjustment for multiple confounders).	At 26 months, infants identified before 6 months of age had better adjusted mean MCDI scores for expressive language (76.2 vs 56.6, $p=0.001$), receptive language (82.1 vs 58.3, $p=0.002$), MacArthur CDI adjusted mean receptive vocabulary (200 vs 86.4, $p<0.001$), expressive vocabulary (117 vs 54, $p<0.03$)

Table 3. Cohort studies reporting language outcomes (cont.)

Study, Year	Selection of Subjects	Comparability and Maintenance of Early vs. Late Groups	Adjustment for Confounders	Results
Yoshinaga-Itano, 1998[88]	Convenience sample of 150 children 13 to 36 months of age with mild to profound SNHL, divided into those identified before (n=72) or after (n=78) 6 months of age. The number of low-risk infants and the role of UNHS in identifying subjects are not described. Selection bias is likely because the design probably excluded infants diagnosed to have hearing loss but did not enter the program, or entered, but were lost to followup.	At baseline, compared groups differed in some demographic characteristics and in the proportion of subjects with cognitive impairment and severe & profound HL.(CQ <80, 29% early group vs. 56%, late group). Severe & profound HL 34% early group vs 46% late group. No report of attrition or followup rates.	There was stratification by cognitive quotient (<80 vs. >80). Other covariates (gender, minority status, maternal education level, Medicaid status, severity, mode of communication, other disabilities) were examined singly in 2-way ANCOVAs.	At 13-36 months, adjusted mean MCDI receptive language quotient (LQ) was higher for those identified before 6 months (79.6 vs. 64.6, p<0.001). Mean MCDI expressive LQ was higher (78.3 vs 63.1, p<0.001) as well as total language (79 vs 64, p<0.001) higher in early identified group . No difference in LQs between 4 age of identification levels in late group.
Mayne, 2000[92]	Convenience sample of 113 children 24-37 months of age, divided into those diagnosed before and after 6 months of age. The number of low-risk infants and the role of UNHS in identifying subjects are not described. Overlap of sample with previous CHIP studies was not reported.	Demographic comparisons of the groups were not reported. No report of attrition or followup rates.	Regression analysis adjusted for degree of hearing loss, mode of communication, other disabilities, parents' hearing, cognitive quotient, mother's education, ethnicity, SES	At 24-36 months, age at diagnosis explained 23% of the variance in expressive language scores.

Table 3. Cohort studies reporting language outcomes (cont.)

Study, Year	Selection of Subjects	Comparability and Maintenance of Early vs. Late Groups	Adjustment for Confounders	Results
Studies from other programs				
Moeller, 2000[91]	Convenience sample of 112 5 year olds who completed the Diagnostic Early Intervention Program in Lincoln, Nebraska. Children with non-verbal IQ <70 and those who did not participate in program through age 5 were excluded. The number of low-risk infants and the role of UNHS in identifying subjects are not described. Outcome assessments were made pre- and post-intervention.	Not reported. No report of attrition or followup rates. Early identified children may have more opportunity to drop out, although differential drop out may be less of a problem at 5 years than in studies assessing closer to enrollment.	Multiple regression analysis adjusted for family involvement, degree of hearing loss and non-verbal IQ.	At age 5, family involvement accounted for 57% of variance in vocabulary and age of enrollment accounted for 11.5%. Adjusted mean vocabulary and reasoning scores were within normal range among children enrolled prior to 11 months but were lower for later-identified children.
Calderon, 2000[93]	Cohort of 80 children with profound hearing loss enrolled in Early Child Hearing Intervention (ECHI) in Seattle Washington. Children with developmental delay were excluded. Cohort grouped by 3 levels by age of entry into program: <1 year (n=9), 12-24 months (n=39), >24 months (n=32). The method of sampling is not described, but the design excluded patients who entered the program but did not graduate.	Not reported. Late diagnosed group had less severe-to-profound loss (36% vs. 66%). Overall loss to followup not reported. Because the early-diagnosed group were in the program longer, they had more opportunity to drop out, so a differential loss to followup is likely.	Controlled for degree of hearing loss, degree of outcome impairment that was present upon entry into program (baseline test levels).	At 3 years, age at entry to program explained 43.5% of the variance in receptive language and 49% of the variance in expressive language. Children treated before 2 years had better outcomes than those treated after 2 years. Only 3 children entered the program prior to 6 months of age.

Table 4. Benefits of screening a hypothetical cohort of 10,000 newborns for moderate to profound PHL

Benefit and Relevant Factors	Probability or effect size	UNHS	High-risk screening
Assumptions*			
Proportion high risk	0.2		
Prevalence			
High risk group	0.008		
Low risk group	0.0008		
Miss rate for UNHS (proportion not screened in hospital)			
in high risk	0.1		
in low risk	0.05		
followup rate for misses	0.9		
Miss rate for high-risk screening			
in high risk^	0.2		
followup rate for misses	0.75		
Sensitivity of 2-stage screening	0.85		
Specificity of 2-stage screening	0.97		
Compliance with followup	0.9		
Accuracy of diagnostic ABR			
Sensitivity^	1		
Specificity	0.995		
Proportion of low-risk diagnosed before 10 months without screening^	0.35		
Treated before 1 year	0.6		
Results			
Number of infants screened		9400	1600
Cases diagnosed before 10 months		17	12
Cases treated before 10 months		10	7
out of total		*22*	*22*
"False Positive" screening tests		254	48
Normal infants incorrectly diagnosed to have PHL at first post-hospital audiologic examination		1	0
NNS to diagnose 1 case		584	173
NNS to diagnose 1 additional case before 10 months		1441	
NNS to treat one additional case before 10 months		2401	

PHL indicates permanent hearing loss; NNS, number needed to screen; ABR, automated brainstem response
UNHS=universal newborn hearing screening

* Base case assumptions are derived from the studies in Table 2, except for those marked with ^.

Table 5. Strength of evidence for universal newborn hearing screening.

Key Question	Evidence Code	Quality of Evidence
1. Can UNHS accurately diagnose moderate to severe sensorineural hearing impairment?		
Does UNHS improve the yield of screening, compared with selective screening of high-risk newborns? *In UNHS increases identification of deaf and hearing-impaired infants between 18.5% and 33% over selective screening in high-risk children.*	II-1, II-3	Good: Consistent results from several cohort studies that recorded risk factors to estimate the proportion of cases that had no risk factors. However, no controlled trials of UNHS versus selective screening have been done.
What are the sensitivity, specificity, and predictive value of screening tests? *OAE and ABR are highly accurate screening tests for congenital SNHL and are clearly more accurate than home visitor screening (using the response to noise as a screening test at ages 6 to 12 months).*	II-1	Good: One controlled trial measured the predictive value of a positive test result, 6.7%, and a good quality cohort study measured sensitivity and specificity against an independent gold standard.
In screening programs, how many children are identified and treated before six months? *UNHS increases the chance that diagnosis and treatment will occur before 6 months of age.*	II-1, II-3	Fair: One controlled study in the United Kingdom and one cohort study in the U.S. reported the frequency of treatment before 10 and 5 months, respectively. Other studies did not provide sufficient information, and none included patients who, although screened, were diagnosed and treated late because of loss to followup.
2. Does screening improve language and communication skills? *Evidence is inconclusive.*	II-2	Poor: One matched, retrospective study with selection bias and other flaws. No well-conducted controlled studies that compare language outcomes in children identified by UNHS to those identified by usual care or selective screening.
3. Does identification and treatment prior to 6 months improve language and communication in infants who would not be diagnosed that early in a selective, high-risk screening program? *Evidence is inconclusive.*	II-2, II-3	Fair/Poor: Studies suffer from selection bias and baseline differences between compared groups. These studies did not specifically describe outcomes in the subgroup of children who would be identified by UNHS but not by selective screening.
4. What are the potential adverse effects of screening and of early treatment?	III	Poor: Most postulated adverse effects have not been evaluated in studies.

Evidence codes: I, Randomized, controlled trial; II-1, Controlled trial without randomization; II-2, Cohort or case-control analytic study; II-3, Multiple time series, dramatic uncontrolled experiments; III, Opinions of respected authorities

Evidence Table 1. Studies reporting speech and language outcomes

Study, Year, Location	Study Design	Population	Exposure & Treatment	Confounders	Outcomes	Age at Testing	Results	Comments
Yoshinaga-Itano, 2000[87] Colorado	Retrospective study of children enrolled in the Colorado Home Intervention Program (CHIP)	25 matched pairs of children enrolled in CHIP program since 1996	UNHS program. Children born in a hospital with a UNHS program in effect at time of birth were compared to children born in in a hospital without a UNHS program in effect at time of birth	Children matched on degree of hearing loss (mild moderate, moderately-severe, profound), cognitive quotient and age at time of speech & language evaluations. No confounding by gender, ethnicity, presence of multiple disabilities, mode of communication, education of primary caregiver, chronological age. All children enrolled in CHIP program for treatment and therapy. This is a program that emphasizes parental involvement	Receptive, expressive and total language skills. Minnesota Child development Inventory (MCDI) used to measure expressive language and comprehension-conceptual (receptive language). Parent report measure used. MacArthur Communicative Development Inventory used to measure vocabulary words produced by child with either spoken or sign language. Videotape of parent/child interaction (25 minutes). Language quotient (LQ) calculated by dividing the child's age score by chronological age and multiplying by 100. Total LQ is an average of receptive and expressive LQ scores. Scores 80 and above are considered normal language scores, 70-80 is borderline normal and 70 and below indicates language delay.	9-59 months; pairs matched on age of testing	Mean scores for expressive, receptive and total language within normal range for screened group and 18-21 points higher ($p<.001$) than unscreened group whose scores indicated language delay. Mean scores, screened compared to unscreened were: expressive language 82.9 (SE 3.7) vs 62.1 (SE 4.3), receptive language-81.5 (SE 3.7) vs 66.8 (SE 4.0); total language-82.2 (SE 3.3) vs 64.4 (SE 3.9). Discrepancy quotient scores (DQs) were significantly lower in screened group (14.8 vs 34.5), $p<.001$. DQ evaluated by both screening and age-of-identification categories.	Study results indicate children exposed to UNHS have significantly better speech and language development than children who do not have the opportunity of screening at birth. Expressive vocabulary for the 75th percentile of the screened group was ten times greater (500 vs 50 words) than the 75th percentile of the non-screened detected group. Differences in language outcomes appear to be due to early identification of hearing loss. Results in agreement with the larger cohort study of Yoshinaga-Itano, 1998. Children matched on most important contributors to language development.

Evidence Table 1. Studies reporting speech and language outcomes

Study, Year, Location	Study Design	Population	Exposure & Treatment	Confounders	Outcomes	Age at Testing	Results	Comments
Yoshinaga-Itano, 2000[87] Colorado (cont'd.)					Discrepancy quotient scores (DQs) calculated to measure children's language relative to cognitive ability. DQ=CQ-total LQ. A high DQ score indicates a large lag in language compared to a population without hearing loss.		Children identified prior to 6 months (whether in screened or unscreened group) had significantly lower DQ scores indicating a smaller gap between language development and cognitive ability. Language development (total LQ) was within normal range for 56% of screened group compared to 24% of unscreened group. Delayed language development (<70) was present in 24% of the screened group compared to 68% of unscreened group.	no confounding by other factors. Cases and controlls enrolled in CHIP program which is parent centered and uniform throughout state.

Evidence Table 1. Studies reporting speech and language outcomes

Study, Year, Location	Study Design	Population	Exposure & Treatment	Confounders	Outcomes	Age at Testing	Results	Comments
Yoshinaga-Itano, 1998[88] Colorado	Retrospective cohort study Convenience sample from population-based program (CHIP)	n=150; ages 13 mos-3 yrs 72 ≤ 6 mos 78>6 mos; severe & profound loss 34% early identified gp & 46% late identified group	Identification & treatment of HL prior to 6 mos of age; Colorado Home Intervention Program (CHIP)	Gender, ethnicity, mother's education Medicaid status, degree of HL, mode of communication, multiple handicaps age at data collection, cognitive ability CQ <80 vs ≥ 80 CQ=Play Assessment Qx age score/ chronological age x 100. Cognitive status measured using tests that are not influenced by hearing or language.	Receptive and expressive language skills. Minnesota Child Development Inventory (MCDI). Calculate language quotient (LQ) for receptive expressive & total language. LQ=100 if chronlogical age same as language level	Age at data collection 13 36 months. Time between identification of HL and hearing aid fitting 2 months for early group (gp 1) and 1 month for late group (gp 2)	Receptive LQ, expressive LQ, total LQ higher in early identified group compared to group identified > 6mos (ANCOVA p>0.001). Adjusted mean LQs 79.6 (sd=25.8) receptive language, 78.3 (sd=26.8) expressive language, 79.0 (sd=25.6) early group compared to later identified group LQs adjusted mean LQs 64.6(sd=20.9) receptive language, 63.1 (sd=19.8) expressive language, 63.8 (sd=19.3). No difference in LQs between 4 age of identification levels in late group. Number of subgroup analyses performed to evaluate the effect of demographic variables.	Best study found which evaluates the effect of early identification and treatment on language skills. Appropriate statistical measures used to adjust for confounding variables, particularly CQ. Source of cases not described- What proportion of HOH children in age range 13 mos-3 years does this represent? Screening and detection program not described. Colorado has had HRR screening for a number of years and more recently added UNHS. No information on premature delivery in either group. Severity of hearing loss greater in late identified group. No trend test.

Evidence Table 1. Studies reporting speech and language outcomes

Study, Year, Location	Study Design	Population	Exposure & Treatment	Confounders	Outcomes	Age at Testing	Results	Comments
Yoshinaga-Itano, 1998[27] Colorado	Retrospective Cohort study Convenience sample from population based program (CHIP)	n=82; age of identification 34 ≤ 6 mos 48 > 7-18 mos; early gp 42% severe/ profound loss vs 67% severe/ profound loss late gp	Identification & treatment of HL prior to 6 mos of age; CHIP	Gender, degree of HL, cognitive function, age at identification, other disabilities (30%), DQ< 60 excluded SES-those from HRR predominately medicaid	Receptive and expressive language skills. Minnesota Child Development Inventory (MCDI); MacArthur Communication Development Inventory (CDI)	Age at data collection; early gp 19-36 months mean age 26.5 (sd=5.9) late gp 27.6 (sd=5.1)	Early gp had higher levels of receptive & expresive language, personal social development, expressive and receptive vocabulary, general development, situational comprehension and vowel production. ANCOVA for age of identification and degree of hearing loss controlled for CQ conducted on dependent measures. Adjusted means for early gp: expressive language 76.2, comprehension-conceptual 82.1, receptive vocab 200, expressive vocal 117 vs 56.6, 58.3, 86.4 and 54.6 for late gp	Early group identified from HRR, source of late gp not specified; presume referred to CHIP by usual care. All children in study selected from CHIP database if there was information on age of identification, degree of hearing loss, cognitive functioning, and chronological age at time of testing. Early group had higher % of children with mild-moderate loss and additional disabilities than did late group; < 6 mo group found mainly via HRR; late gp referred by usual care. Study sample included all children in CHIP database with required information

Evidence Table 1. Studies reporting speech and language outcomes

Study, Year, Location	Study Design	Population	Exposure & Treatment	Confounders	Outcomes	Age at Testing	Results	Comments
Yoshinaga-Itano, 1998[89] Colorado	Retrospective Cohort study Convenience sample from population based program (CHIP)	n=40; age of identification 15 < 6 mos 25 >18 mos ; Early gp 47% severe/ profound hl, late gp 52% severe/ profound HL	Identification & treatment of HL prior to 6 mos of age. Colorado Home Intervention Program (CHIP)	Gender, degree of HL, cognitive function, age at identification, other disabilities (30%), DQ< 60 excluded SES-those from HRR predominately medicaid	Receptive and expressive language skills. Minnesota Child Development Inventory (MCDI); MacArthur Communication Development Inventory	Early gp-mean age 39.9 months sd-11.9; late gp- 40.7 sd-10.7	Adjusted means for expressive language and comprehension-conceptual subtests of MCDI significantly higher for early group. P<.05; Expressive lang 81.1 vs 64.3; comprehension-conceptual 84.4 vs 70.1.Children in early group have 8 month delay expressive lang and 6 mos delay receptive language compared to delay of 14 mos and 12 mos for late group.	Convenience sample of children from CHIP. Presume late identified children referred from usual care to CHIP. Children in both early and late gps similar for age at testing, hearing loss by audiologic category and self help developemental quotient. Differered only on age of identification mean age 1.9 mos vs 26 mos. ANCOVA used; small sample size. Required data on age of ID, age at testing, MCDI scores, degree of HL
Apuzzo 1995[90] Colorado	Retrospective Cohort study Convenience sample from population based program (CHIP)	n=69; age of identification birth-2 mos n=14, 3-12 mos n=11, 13-24 n=30, 25+ mos n=12. 50% early gp severe/ profound loss vs 65% of later identified group	Identification and treatment at 4 different ages; CHIP	Exclude children with DQ <60 to control for cognitive delay; age of testing same for all 4 groups; controlled for hearing loss	Receptive and expressive language skills. Minnesota CDI MacArthur CDI	Age at testing similar for all groups, about 40 months	Early identified group(0-2 mos) scored higher on general development and expressive language subtests of MCDI than did children identified later, p<0.01 One way ANOVA used	Group identified earliest were all from HRR screening; others referred to CHIP from usual care. Small sample size. Only 2 children w profound loss in early gp. No adjustment for SES

Evidence Table 1. Studies reporting speech and language outcomes

Study, Year, Location	Study Design	Population	Exposure & Treatment	Confounders	Outcomes	Age at Testing	Results	Comments
Moeller, 2000[91] Nebraska	Retrospective cohort study	n=112 graduates of Diagnostic Early Intervention Program (DEIP)	Received intervention through age 5. Identified by DEIP; 61 in auditory-oral program; 51 in total communica-tion program; hearing aids, FM trainers and appropriate speech/lang-uage therapy	Exclude children with non-verbal IQ <70 and those who did not participate in program through age 5; controlled for family involvement, degree of hearing loss and non-verbal IQ	Peabody Picture Vocab Test (PPVT), preschool language assessment instrument. Developed a 1-5 scale to rate degree of family involvement in child's educational program	age 5 ; tested at end of preschool period, at entry to kindergar-ten	Multiple linear regression used. Age of enrollment and degree of family involvement significantly correlated with higher vocabulary scores controlling for IQ and degree of hearing loss. Family involvement accounted for 57% of variance in vocab scores and age of enrollment an additional 11.5% Children enrolled prior to 11 months had stronger vocab and reasoning skills than later-enrolled children; early enrolled children within range of hearing peers. Vocab scores decreased as age of enrollment increased (≤ 11mos, 11.1-23 mos, 23.1 35 mos, >35mos)	Children received periodic hearing and language evaluations during preschool intervention program. 110/112 (98%) in structured program 67/112 (59.8%) had severe (n=20) or profound (n=47) hearing loss.

Evidence Table 1. Studies reporting speech and language outcomes

Study, Year, Location	Study Design	Population	Exposure & Treatment	Confounders	Outcomes	Age at Testing	Results	Comments
Mayne, 2000[92] Colorado	Cohort study	113 children enrolled in CHIP program. Ages 24-37 months	CHIP intervention described in previous studies	No factors found to be associated with language development in deaf. Had information on degree of hearing loss, mode of communication, age of identification, presence of other handicaps, parents hearing status, IQ as measured by cognitive quotient (CQ), mother's education, ethnicity, SES, CQ>-80 vs CQ<80.	Expressive vocabulary as measured by MacArthur Communicative Development Inventory	Continuing evaluation as part of the CHIP program	Bivariate regression models. Full model explained 56% of variance in expressive vocabulary scores. Expressive vocabulary higher with increased age, increased CQs, identification of hearing loss by the age of 6 months and having a hearing loss as the only medical condition. Normative data presented for deaf and hard-of-hearing children, ages 8-37 months.	Study identified factors contributing to variations in expressive vocabulary scores in a group of deaf and hard-of-hearing children 24-37 months. Effects of gender and SES not determined. However, these variables had no effect in this study. Excellent literature review of language development in children with normal hearing and review of language development and school performance in the deaf and hard-of-hearing.

Evidence Table 1. Studies reporting speech and language outcomes

Study, Year, Location	Study Design	Population	Exposure & Treatment	Confounders	Outcomes	Age at Testing	Results	Comments
Calderon, 2000[93] Washington	Retrospective cohort study	80 children enrolled in Early Child Hearing Intervention (ECHI) program, a home based total communica-tion program in western Washington. Cohort grouped by 3 levels by age of entry into program: <1 year (n=9), 12-24 months (n=39), >24 months (n=32)	Early identification and treatment. All children had profound hearing loss and all had been fitted with hearing aid or cochlear inmplants.	Controlled for degree of hearing loss, degree of outcome impairment that was present upon entry into program (baseline test levels). Excluded children with developmental delay.	Receptive and expressive language development, pre- and post-test auditory development and speech production skills. Evaluated at 3 years, at end of program.	Entry to program and at 36 months. Pre and post measures used	ANOVA and multiple regression used. Age at entry to program explained 40% of the variance within the data. Significant trend for age at intervention, suggesting that children with an intervention at < 1 year had better outcomes than children with interventions at 12-24 months, who in turn had better outcomes than children starting even later, at 24 months. No benefit found for age at identification and speech and auditory scores.	All children enrolled in same program with continuous records and testing. Pre/post measures same for all in group. Review of literature on normal language development landmarks.

Evidence Table 2. Quality rating of cohort studies reporting language outcomes

Study, Year (Quality)	Selection of Subjects	Comparability and Maintenance of Early vs Late Groups	Adjustment for Confounders	Results
Studies from Colorado Home Intervention Program (CHIP)				
Apuzzo, 1995[90] (FAIR)	Convenience sample of 69 high-risk infants diagnosed between 2 and 25 months of age. Children with severe cognitive delay were excluded.	Late-identified group was more likely to have severe to profound hearing loss (65% vs 50%). No report of attrition or followup rates.	One-way ANOVA did not adjust for SES, family involvement, or other potential confounders.	At 40 months of age, infants identified before 2 months of age had higher mean Minnesota Child Development Inventory (MCDI) scores for expressive language ($p<0.01$).
Yoshinaga-Itano, 1998[89] (POOR)	Convenience sample of 40 high-risk infants, divided into those identified and treated before 6 months of age (n=15) and those treated after 18 months (n=25). Children with severe cognitive delay were excluded (DQ<60).	Late-identified group was more likely to have severe to profound hearing loss (52% vs 47%). No report of attrition or followup rates.	Gender, severity of hearing loss, cognitive function, and other disabilities were examined in 2-way ANCOVAs, not in a multiple regression (no simultaneous adjustment for multiple confounders).	At 40 months, infants identified before 6 months of age had better adjusted mean MCDI scores for expressive language (81.1 vs 64.3, $p<0.05$) and receptive language (84.4 vs 70.1, $p<0.05$).
Yoshinaga-Itano, 1998[27] (POOR)	Convenience sample of 82 infants, 19 to 36 months of age, with mild to profound PHL, divided into those identified before 6 months of age (n=34) and between 7 and 18 months of age (n=48). Early group identified by high-risk registry; late group by usual care. Children with severe cognitive delay were excluded (DQ<60).	Late-identified group was more likely to have severe to profound hearing loss (77% vs 42%). No report of attrition or followup rates.	Gender, severity of hearing loss, cognitive function, and other disabilities were examined in 2-way ANCOVAs, not in a multiple regression (no simultaneous adjustment for multiple confounders).	At 26 months, infants identified before 6 months of age had better adjusted mean MCDI scores for expressive language (76.2 vs 56.6, $p=0.001$), receptive language (82.1 vs 58.3, $p=0.002$), MacArthur CDI adjusted mean receptive vocabulary (200 vs 86.4, $p<0.001$), and expressive vocabulary (117 vs 54, $p<0.03$).

ANOVA indicates analysis of variance; SES, socioeconomic status; DQ, developmental quotient; ANCOVA, analysis of covariance; PHL, permanent hearing loss; UNHS, universal newborn hearing screening; CQ, cognitive quotient; LQ, language quotient; SE, standard error

Evidence Table 2. Quality ratings of cohort studies reporting language outcomes (cont.)

Study, Year (Quality)	Selection of Subjects	Comparability and Maintenance of Early vs Late Groups	Adjustment for Confounders	Results
Yoshinaga-Itano, 1998[88] (POOR)	Convenience sample of 150 children 13 to 36 months of age with mild to profound PHL, divided into those identified before (n=72) or after (n=78) 6 months of age. The number of low-risk infants and the role of UNHS in identifying subjects are not described. Selection bias is likely because the design probably excluded infants who were diagnosed to have hearing loss but did not enter the program, or who entered, but were lost to followup.	At baseline, compared groups differed in some demographic characteristics and in the proportion of subjects with cognitive impairment and severe to profound hearing loss (CQ <80, 29% early group vs 56% late group; severe to profound hearing loss 34% early group vs 46% late group). No report of attrition or followup rates.	There was stratification by CQ (<80 vs >80). Other covariates (gender, minority status, maternal education level, Medicaid status, severity, mode of communication, other disabilities) were examined singly in 2-way ANCOVAs.	At 13 to 36 months, adjusted mean MCDI receptive language LQ was higher for those identified before 6 months (79.6 vs 64.6, p<0.001). Mean MCDI expressive LQ was higher (78.3 vs 63.1, p<0.001) and total language (79 vs 64, p<0.001) was higher in early-identified group. No differences in LQ among 4 age of identification levels in late-identified group.

ANOVA indicates analysis of variance; SES, socioeconomic status; DQ, developmental quotient; ANCOVA, analysis of covariance; PHL, permanent hearing loss; UNHS, universal newborn hearing screening; CQ, cognitive quotient; LQ, language quotient; SE, standard error

Evidence Table 2. Quality ratings of cohort studies reporting language outcomes (cont.)

Study, Year (Quality)	Selection of Subjects	Comparability and Maintenance of Early vs Late Groups	Adjustment for Confounders	Results
Mayne, 2000[92] (POOR)	Convenience sample of 113 children 24 to 73 months of age, divided into those diagnosed before and after 6 months of age. The number of low-risk infants and the role of UNHS in identifying subjects are not described. Overlap of sample with previous CHIP studies was not reported.	Demographic comparisons of the groups were not reported. No report of attrition of followup rates.	Regression analysis adjusted for degree of hearing loss, mode of communication, other disabilities, parents' hearing, cognitive quotient, mother's education, ethnicity, SES	At 24 to 36 months, age at diagnosis explained 23% of the variance in expressive language scores.
Studies from other programs				
Moeller, 2000[91] (FAIR)	Convenience sample of 112 5-year-olds who completed the Diagnostic Early Intervention Program in Lincoln, Nebraska. Children with non-verbal IQ <70 and those who did not participate in program through age 5 were excluded. The number of low-risk infants and the role of UNHS in identifying subjects are not described. Outcome assessments were made pre- and post-intervention.	Not reported. No report of attrition or followup rates. Early identified children may have more opportunity to drop out, although differential drop out may be less of a problem at 5 years than in studies assessing closer to enrollment.	Multiple regression analysis adjusted for family involvement, degree of hearing loss and non-verbal IQ.	At age 5, family involvement accounted for 57% of variance in vocabulary and age of enrollment accounted for 11.5%. Adjusted mean vocabulary and reasoning scores were within normal range among children enrolled prior to 11 months but were lower for later-identified children.
Calderon, 2000[93] (FAIR)	Cohort of 80 children with profound hearing loss enrolled in Early Child Hearing Intervention (ECHI) in Seattle Washington. Children with developmental delay were excluded. Cohort grouped by 3 levels by age of entry into program: <1 year (n=9), 12-24 months (n=39), >24 months (n=32). The method of sampling is not described, but the design excluded patients who entered the program but did not graduate.	Not reported. Late diagnosed group had less severe-to-profound loss (36% vs 66%). Overall loss to followup not reported. Because the early-diagnosed group were in the program longer, they had more opportunity to drop out, so a differential loss to followup is likely.	Controlled for degree of hearing loss, degree of outcome impairment that was present upon entry into program (baseline test levels).	At 3 years, age at entry to program explained 43.5% of the variance in receptive language and 49% of the variance in expressive language. Children treated before 2 years had better outcomes than those treated after 2 years. Only 3 children entered the program prior to 6 months of age.

ANOVA indicates analysis of variance; SES, socioeconomic status; DQ, developmental quotient; ANCOVA, analysis of covariance; PHL, permanent hearing loss; UNHS, universal newborn hearing screening; CQ, cognitive quotient; LQ, language quotient; SE, standard error