Donald Berry, PhD
Chairman, Department of Biostatistics
& Frank T. McGraw Memorial Chair of Cancer Research
University of Texas M.D. Anderson Cancer Center
e-mail: dberry@mdanderson.org

February 28, 2002
Mammography hearing
U.S. Senate

Evaluating the Evidence of Benefit for Screening Mammography

I serve on the PDQ (Physicians' Data Query) Screening and Prevention Editorial Board. We write statements for the NCI Website http://www.cancer.gov/cancer_information/ regarding screening for cancer and preventing cancer. However, we are independent of the NCI. Our statements are intended for and are accessible by physicians and the general public. We meet approximately six times per year to discuss recently published literature and on the basis of the available information we decide whether and how to modify our Website statements. We assign levels of evidence to our statements. Contrary to reports in the press, we are not advisory to the NCI, we do not establish guidelines, and we do not make official recommendations.

I will give my understanding of the discussions and intentions of the PDQ Board. However, I have not been elected to be a spokesperson for the Board and so I do not have the right to speak for other members of the Board.

My introduction to today's topic was my appointment five years ago to an NIH Consensus Development Conference Panel on Breast Cancer Screening for Women Ages 40-49. I had no ax to grind then and I have none now. My life is dedicated to understanding and fighting cancer—breast cancer in particular. I am intimately involved in the prevention and treatment of this horrible disease. Nothing would please me more—professionally and personally—than to have a tool that eliminates breast cancer or that turns it from a disease that kills into one that is chronic but can be controlled.

The randomized trials

At the January 2002 PDQ Board meeting we considered an article authored by Drs. Ole Olsen and Peter Gotzsche of the Nordic Cochrane Collaborative and that appeared in The Lancet in October 2001. This article critiqued the randomized trials that have been conducted to evaluate the benefits of screening mammography and cited a number of deficiencies and flaws. Many of these were known previously and there was little original information in the review. However, it served to put the trials' deficiencies into perspective and led us to re-evaluate the credibility of the trials. We decided to revise our breast cancer screening statement and to refer to the Olsen-Gotzsche article. The plan is to discuss and possibly finalize the revision at our meeting in March. The current version of the statement indicates that the estimates of the benefits of screening are uncertain. Therefore, in a sense the revision will be minor. However, we plan to indicate that the existence of benefit is itself uncertain.

Olsen and Gotzsche reviewed the seven randomized trials. One was conducted in Canada, one in New York, one in Edinburgh, Scotland and the other four in Sweden. The PDQ panel discounted some of the deficiencies pointed out by Olsen and Gotzsche but we agreed with others. In the first category, most of us (1) felt that their focus on all-cause mortality (rather than breast-cancer specific mortality) was too strong, (2) that imbalances in randomization were not a major concern (except in Edinburgh) and (3)

regard the use of mammograms in the control groups (to coincide with the end of the screening period) of three of the Swedish trials to be a reasonable design strategy. From our perspective the trials had four types of major deficiencies. They applied to some but not all of the trials. The first three are potential sources of bias favoring the screening group and in each case there is some evidence of actual bias in the trials.

(1) Women with pre-existing breast cancer were preferentially excluded from the screening group. The problem was most severe in the New York trial in which 853 women in the screened group and 336 in the control group were excluded because they had breast cancer at the time of randomization. Excluding women with breast cancer is not unreasonable, but the numbers excluded in the two groups would be about the same had there been no bias. If these women had been included and only 9% of the differential of 517 women died of their disease, the breast cancer mortality rate would have been higher in the screened group than in the control group.
(2) Attribution of cause of death was made with knowledge of whether the woman was in the screened group. Blinding assessment of cause of death to assigned intervention is fundamental in good clinical trial practice. For example, an assessor might be more likely to attribute a death to lung cancer if the woman's cancer was detected through screening and to metastatic breast cancer if the woman had been in the control group. There is evidence that this bias was real. The numbers of deaths have changed in unusual ways from one report of the trial results to the next: The number of breast cancer deaths in the control group always increases over time but it sometimes decreases in the screened group.
(3) In three of the Swedish trials women in the control group were supposed to have a mammogram, which was scheduled at the time of the last mammogram in the screened group. Then, deaths due to breast cancer in the control group would be counted only if they were diagnosed at or before this mammogram and in the screened group if they were diagnosed at or before the last mammogram. This design is reasonable. But the scheduled control mammogram slipped in all three trials, allowing for more time to detect cancers in the control group. The slippage was by as much as 18 months. As a consequence, the control group in the Göteborg trial had 21% more breast cancers detected than did the screened group. Such an observation seems impossible (in an unbiased design) because mammography is very good at finding breast cancers.
(4) No independent audit of trial results. Having an independent audit is a generally accepted in medical research and it is essential for a trial to be credible. For example, the FDA routinely audits clinical trials that provide the basis for an experimental drug's safety and efficacy. None of the Swedish investigators have opened their results to external inspection (but some have recently indicated their willingness to do so).

The Canadian trial was subject to none of these biases. It has been extensively audited and its data are openly available for external examination. Both parts of the Canadian trial (one admitted women in their 40s and the other admitted women in their 50s) found a higher breast cancer mortality rate in the screened group, although the increase was not statistically significant. The other trials fell prey to one or more of the biases, although it is not known whether there were biases in the first part of the Malmö trial.

How can people differ so in their evaluation of evidence?

Physicians learn by experience. At least 90 percent of what is known in medicine today is the result of clinical observation, with the remaining knowledge deriving from randomized clinical trials. Experience is a great teacher. But when it comes to inferring the benefits of screening, clinical observation is fundamentally subject to flawed interpretation.

Women with breast cancer detected mammographically have extremely good prognoses in comparison with those having cancers detected in any other way. Mammographically detected tumors are smaller and are less likely to have spread to the axillary lymph nodes. Since women whose breast cancers were found by a mammogram do so much better, there is a tendency to attribute the benefit to mammography. Unfortunately, this logic is wrong. The fallacious aspect is not simply a nuance—it is a mistake that gives rise to profound misconceptions. And it is a logical lapse to which doctors and patients alike can fall prey.

Suppose temporarily that screening mammography has no survival benefit. Clinicians would still see precisely what they do see. Consider a 50-year-old woman who has breast cancer and who is destined to die of her disease at age 60. However, she does not yet know that she has breast cancer. It would be found on a mammogram if she were to have one, and she would live for ten years with breast cancer. But without a mammogram it would show up clinically only when its symptoms become apparent, say at age 55. So without a mammogram she lives for only five years after her cancer is discovered. The discrepancy between ten years and five years results from what is called *lead-time bias*. It means that women whose cancers detected by mammography live longer than do those detected otherwise, and this is true even if screening has no true benefit.

There is another kind of bias—called *length bias*—that is even more important in magnitude, but it is not as easy to understand. It is related to the fact that breast cancer is a heterogeneous disease. Again, assume temporarily that screening has no survival benefit. We understand some of the factors that give rise to this heterogeneity, but not all of them. Some cancers grow rapidly and others take a more indolent course. Suppose just for the sake of discussion that there are two kinds of cancers: half grow fast and the other half grow slowly. We cannot determine which is which and so we treat them similarly. Suppose that after their cancer is detected via mammography, patients having the first type live an average of five years and patients with the second type live an average of 35 years (not counting causes of death other than breast cancer). So the average survival for women whose cancers are detected by mammograms is about 20 years. In the absence of mammography the first type of cancer might show symptoms with only three more years to live (a lead-time of two years). Some portion—say one half—of the women who harbor the slowly growing tumors will die of other causes before it is discovered. The other half of these women will discover them with 24 more years to live, say, a lead-time of 11 years. There will be 25 percent fewer breast cancers in the non-mammography group. Two-thirds will live an average of three years and one-third will live an average of about 24 years, for an overall average of ten years. So women diagnosed with mammography live about ten years longer than those detected otherwise. This enormous difference is pure artifact since we assumed that screening had no benefit.

The above assumptions were simplified to make a point. No one thinks that there are only two kinds of breast cancer. But everyone recognizes that the disease is heterogeneous. Length bias and lead-time bias are present regardless of the form of heterogeneity. Together they account for enormous differences in apparent survival, as measured from the date of diagnosis, between screened and unscreened cancer patients. These differences are so large that they are detectable by physicians in their everyday practices.

No wonder physicians are persuaded of screening's benefits. But the observed benefits may be completely spurious. In other words, apparent survival from diagnosis may be longer, but life expectancy may not change at all. Hence the need for randomized trials.

Relative risk vs. absolute risk

If there is a benefit of screening then the benefit is modest. To see this, ignore the criticisms of the trials and take their results at face value. The benefits evinced vary considerably from one trial to the next. Outside of the Canadian trial (which showed no benefit), the highest quality results are from the Swedish trials. The most recent results (out to 18 years) of the Swedish trials show a reduction in breast cancer mortality of 21% (over all ages) in favor of screening. The value 21% is a *relative risk* reduction, which is convenient as a statistical measure of benefit. But relative risk is difficult to interpret clinically. One measure of *absolute risk* is to convert the 21% into expected life gained per woman screened. In the first 18 years following initiation of screening in the Swedish trials the average gain is about 4 days. (In contrast, quitting smoking adds years to one's expected lifetime.) Of course, only those women who are eventually diagnosed with breast cancer share in any benefit. Suppose 10% of the women get breast cancer eventually. Then each woman with cancer gains an average of about 40 days. How this is apportioned among the women diagnosed with cancer is not clear. From the trial results it is impossible to distinguish whether (i) each breast cancer patient gains exactly 40 days, (ii) fewer than one percent of patients gain 18 years or more and the rest gain nothing, or (iii) something between these two extremes. Put another way, it is not possible to know whether a small proportion of lives are saved by screening or a large number of women have their lives extended by a small amount, or some combination of the two.

What should we tell women?

The short answer is "The truth." The benefits of screening are uncertain and women should be told this. They may be confused. Confusion is a legitimate state of knowledge, one that may be appropriate in this case. It is a mistake and it is patronizing to women to pretend that we know something we do not. Women have a right to hear about the risks of screening and about the uncertainties regarding the benefits of screening. They should hear all points of view and then decide for themselves. Making this decision will not be easy for some women. We should provide them with decision aids that will inform them of what is known and help them weigh the benefits and risks.

The risks of screening may seem minor but they are important nonetheless, and they are common. From four percent to ten percent of women screened are found to have an abnormal result. The ensuing recommendations range from a follow-up mammogram to having a biopsy. Eighty to 95 percent of the abnormalities turn out to be benign. Obviously, not having cancer is good news, but an estimated 28 million women have mammograms each year, and so a million or more go through the anxious experience of an abnormal test until the final result is known. After ten mammograms the cumulative risk of a false positive result is about 50 percent and about 1 in 6 have biopsies that turn out to be negative. In addition, we know that screening misses about 15 to 25 percent of breast cancers.

Another potential consequence is overdiagnosis. Some breast cancers that may never have progressed become symptomatic during a patient's lifetime. We don't know which of these cancers will progress and so essentially all women with screening-detected breast cancer are treated surgically, with or

without radiation. This may result in unnecessary surgery for some women. Of course, even this serious consequence may be acceptable if the test is saving the lives of other women.

A problem with setting guidelines such as those we have now is that it conveys the message to physicians that screening is an imperative health measure. A woman who decides that the risks outweigh the benefits should not be made to feel that her decision is somehow irrational. A 58-year-old woman from New Jersey sent me the following lament: "Sadly, in my experience anyway, I have found it impossible to have a rational conversation with a physician, where my concerns are respected on the topic of mammograms, as the NYTimes article says a patient should have. Doctors get belligerent and almost hostile if I say I have reservations about getting a yearly mammogram. The upshot is that I don't feel I have a good relationship with a physician, and that is not good. A good scientist is not afraid to express uncertainty on a topic or to discuss a topic openly. I'm afraid the practicing physicians who I have come across do not have that scientific mind-set."

Where to go from here?

It is not possible to do another randomized trial, at least not in the United States. Women want either to be screened regularly or not. Few would let a coin toss make their decision. However, there are developments that may help elucidate the issue, and steps that we can take.

(1) Provide women with decision aids in which they are informed of the benefits and risks, including uncertainties, and helped to weigh them in making a decision.
(2) Audit of the Swedish trials. A positive consequence of the PDQ's position and the ensuing discussion in the press was reported by John Crewdson in the Chicago Tribune of January 31, 2002: Several of the Swedish investigators "announced last week that they would release their detailed data, including patient files, to researchers at the U.S. National Cancer Institute or another international body." (Hopefully, the recently announced NCI guidelines will not lead to the Swedes withdrawing this offer.) If an audit of these trials examines the biases and confirms the recently announced 21% reduction in breast cancer mortality then I for one will agree that screening has a benefit.
(3) Cancer Intervention and Surveillance Network (CISNET). This is an NCI-sponsored program that considers a variety of cancers. I am one of seven Principal Investigators considering breast cancer. Breast cancer mortality in the United States has decreased by nearly 15% over the last decade. This coincides with the wide scale introduction of screening mammography. It also coincides with the dramatic upsurge in the use of tamoxifen and improvements in chemotherapy. We use statistical modeling to conclude how much screening mammography, hormonal therapy and chemotherapy have contributed to this decrease. Of special interest is the possibility of synergism between screening and treatment. For example, it may be that treatment with tamoxifen and chemotherapy has more benefit when a tumor is discovered by a mammogram at an earlier stage. We use annual data concerning who got screened, who used tamoxifen, etc. An advantage of this approach is that it applies to mammography actually used in practice in the late 1980s and into the 1990s, which may have been better than that used in the randomized trials. Another advantage is that we assess effectiveness in the context of actual clinical practice rather than in the possibly artificial world of clinical trials.
(4) The third development is the most promising of all. Our understanding of the biology of breast cancer has increased greatly in recent years, but we still know relatively little. Breast cancer would

not be fatal if it were to stay in the breast. Its lethality stems from its penchant for traveling to and setting up shop in other places in the body, such as in bone, the lungs, liver and brain. The question is, When does it do these things? Perhaps cancers manifest their metastatic potential (or not) when they are tiny, say when they total only a million or so cells. If so then they will have dispatched their malevolent messengers from the breast to the rest of the body before even the best mammography can detect their presence. Or it may be that they start sloughing off tumor cells only when they become large enough to have been detected and removed. We know little about such matters. And we know little about the relationship between the biological characteristics of tumors and how to treat them. These issues are being addressed by researchers around the world. Research progress will help us better understand the relationships between biological markers, early detection and treatment. Especially exciting are the genomics and bioinformatics revolutions. These are in their infancies and are well funded, but they deserve all the attention they have received.

Thank you for the opportunity to discuss this extremely important issue in women's health, a topic to which I have and will continue to dedicate my career. I would be happy to answer questions or provide further details.