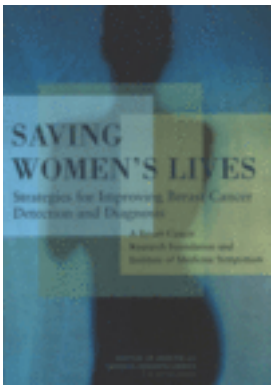

Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis -- A Breast Cancer Research Foundation and Institute of Medicine Symposium

Roger Herdman and Larry Norton, Editors, Committee on New Approaches to Early Detection and Diagnosis of Breast Cancer

ISBN: 0-309-09438-0, 144 pages, 6 x 9, (2005)

SAVING WOMEN'S LIVES

Strategies for Improving Breast Cancer Detection and Diagnosis

A Breast Cancer Research Foundation
and Institute of Medicine Symposium

Committee on New Approaches to Early
Detection and Diagnosis of Breast Cancer

National Cancer Policy Board

Roger Herdman and Larry Norton, *Editors*

INSTITUTE OF MEDICINE *AND*
NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

Support for this project was provided by the Apex Foundation, The Breast Cancer Research Foundation, the Broad Reach Foundation, the Carl J. Herzog Foundation, Mr. Corbin Gwaltney, Mr. John Castle, the Josiah Macy Jr. Foundation, the Kansas Health Foundation, and the National Cancer Institute. This study was supported, in part, by Contract No. N01-OD-4-2139, TO #110 between the National Academy of Sciences and the National Cancer Institute. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number: 0-309-09438-0 (Pbk)
International Standard Book Number: 0-309-54631-1 (PDF)

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

For more information about the Institute of Medicine, visit the IOM home page at: **www.iom.edu.**

Printed in the United States of America.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

.

COMMITTEE ON NEW APPROACHES TO EARLY DETECTION AND DIAGNOSIS OF BREAST CANCER

EDWARD E. PENHOET (*Chair*), Director, Science and Higher Education Programs, Gordon and Betty Moore Foundation, San Francisco, CA

DIANA B. PETITTI (*Vice Chair*)**,** Senior Scientific Advisor Health Policy and Medicine, Kaiser Permanente Southern California, Pasadena, CA

MARTIN D. ABELOFF, Director, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University, Baltimore, MD

COLIN B. BEGG, Chairman, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center

M. KATHLEEN BEHRENS, General Partner, RS & Co. Venture Partners IV, L.P. RS Investments, San Francisco, CA

RICHARD BOHMER, Assistant Professor, Harvard Business School, Boston, MA

CHRISTINE BRUNSWICK, Vice President, National Breast Cancer Coalition, Washington, DC

SANJIV S. GAMBHIR, Director, Molecular Imaging Program at Stanford, Stanford University, Palo Alto, CA

ROBERT A. GREENES, Professor, Radiology and Health Sciences & Technology, Harvard Medical School, Brigham and Women's Hospital, Boston, MA

JEFFREY R. MARKS, Associate Professor, Department of Surgery, Duke University Medical Center, Durham, NC

BARBARA J. McNEIL, Ridley Watts Professor and Head, Department of Health Care Policy, Harvard Medical School, Boston, MA; Professor of Radiology, Brigham and Women's Hospital, Boston, MA

DAVID C. MOWERY, Professor, Walter A. Haas School of Business, University of California, Berkeley, CA

ETTA D. PISANO, Chief of Breast Imaging, Department of Radiology, University of North Carolina Hospitals, Chapel Hill, NC

GLENN D. STEELE, President and CEO, Geisinger Health System, Danville, PA

KIRBY G. VOSBURGH, Associate Director, Center for Integration of Medicine and Innovative Technologies (CIMIT), Cambridge, MA

WILLIAM C. WOOD, Chief of Surgery, Emory University Hospital, Atlanta, GA

Consultants

LAURA J. ESSERMAN, Director, Carol Franc Buck Breast Care Center, University of California, San Francisco, CA
LANCE A. LIOTTA, Chief, Laboratory of Pathology, National Cancer Institute, Bethesda, MD
LARRY NORTON, Deputy Physician-in-Chief and Director of Breast Cancer Programs, Memorial Sloan-Kettering Cancer Center, New York, NY

Study Staff

ROGER C. HERDMAN, Director, National Cancer Policy Board
ELIZABETH J. BROWN, Research Assistant
MARY ANN F. PRYOR, Project Assistant
ANIKE L. JOHNSON, Administrative Assistant

LARRY NORTON, Chair, Medical Advisory Board, The Breast Cancer Research Foundation

NATIONAL CANCER POLICY BOARD

Contents

Abstract

In this report The Breast Cancer Research Foundation (BCRF) and the Institute of Medicine (IOM) present a one-day symposium that was held at the IOM to further disseminate the conclusions and recommendations of the joint IOM and National Research Council report, *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis*. The symposium was introduced by Mrs. Evelyn Lauder, Founder and Chairman of the BCRF; and Dr. Edward Penhoet, Chairman of the IOM committee for the report. At a plenary session in the morning six invited experts from academia, the American Cancer Society, and the Centers for Medicare and Medicaid Services gave presentations on what women need to know about mammography, challenges to expanding mammography capacity, better models for mammography services, risk stratification for breast cancer detection, the promise of biomarkers, and bringing new technologies into service. In the afternoon, panelists from the Natioanl Cancer Institue, the Food and Drug Admnistration, Kaiser Permanente, Blue Cross Blue Shield, the American College of Radiology (ACR), Partners Health Care, and University of California, San Francisco, gave presentations and participated in discussions with attendees in two groups, one on delivering better breast cancer screening services and one on developing and delivering new detection technologies. A wrap-up session at the end summarized the issues raised, including: how to organize mammography better and recruit women to screening; lessons learned from screening in the United Kingdom; and disparities in breast cancer screening and care, Mammography Quality Standards Act enforcement, ACR perspectives on screening, the roles of National Institutes of Health in cancer detection technology development, engineers working with clinicians to develop breast cancer detection, technology evaluation and coverage policy, and assessing technologies in managed care systems. At the end, a representative of breast cancer survivors gave her perspectives on the day.

1

Introduction

In 1996, the National Cancer Institute (NCI) and the Centers for Disease Control and Prevention (CDC) discussed with the National Academies' National Research Council and Institute of Medicine (IOM) the advantages of creating a National Cancer Policy Board (the Board) administered by the IOM. In 1997, funded primarily by the NCI and CDC with some private sector contributions (for example, the American Cancer Society), the Board was established as a division of the IOM. The Board has 18 members with expertise and experience in cancer medicine, science, and advocacy drawn from the national cancer community. As an independent entity, the Board sets its own agenda which involves identifying emerging policy issues in the nation's effort to combat cancer and preparing reports that address those issues.

Before the establishment of the Board, the IOM had begun a long history of exploring and reporting the subject of breast cancer, including: *Breast Cancer: Setting Priorities for Effectiveness Research* (Institute of Medicine and National Research Council, 1990), which recommended to the Health Care Financing Administration research directions on subjects like improved databases, variations in breast cancer care and outcomes, and general methodological issues like indices of severity and case-mix; and *A Review of the Department of Defense's Program for Breast Cancer Research* (Institute of Medicine, 1997), which advised the U.S. Army Medical Research and Materiel Command on progress and future directions of its Breast Cancer Research Program.

The Board, after it was established, built on this experience by releasing under its aegis three additional reports: *Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer* (Institute of Medicine and National Research Council, 2001), which reviewed technologies for breast cancer screening and diagnosis that were in various stages of development; *Meeting Psychosocial Needs of Women with Breast Cancer* (Institute of Medicine and National Research Council, 2004), which examined the current state and future needs of psychosocial research and care for women with breast

cancer; and *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis* (Institute of Medicine and National Research Council, 2005), which reviewed issues in breast cancer screening and the development and deploying of new approaches to early detection of breast cancer.The symposium reported here seeks to describe and disseminate the content and recommendations of this last report.Furthermore, this sequence has not yet come to an end. As this symposium was being held, another committee under the aegis of the Board, began work on *Improving Mammography Quality Standards*, a study of selected issues important to the federal Mammography Quality Standards Act that was requested by the U.S. Congress to inform the planned reauthorization of the Act in 2005.

Over the years, five federal agencies have taken the lead in sponsoring the breast cancer work of the IOM and the Board—the NCI, CDC, and Food and Drug Administration recently and earlier the Health Care Financing Administration and the Department of the Army. For the *Mammography and Beyond* and *Saving Women's Lives* projects, a group of foundations, prominently The Breast Cancer Research Foundation (BCRF), the Broad Reach Foundation, and the Apex Foundation,[1] took the lead, accompanied by a number of other private donors, the Carl J. Herzog Foundation, the Josiah Macy, Jr., Foundation, the Jewish Healthcare Foundation, the Kansas Health Foundation, Mr. Corbin Gwaltney, Mr. John Castle, and the New York Community Trust.

Since 1993, The Breast Cancer Research Foundation has been raising funds, now amounting to about $100 million, to support innovative research in preventing and curing breast cancer. The BCRF was a major supporter of the report, *Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis* and also was prepared to contribute to the costs of planning and disseminating the information and urging the actions described in that report. The BCRF and the Board decided that getting out the message of the report could best be accomplished by a symposium assembling and hearing from those who knew most about breast cancer screening and early detection. They were asked to share insights and consider ways in which the objectives of the report could be achieved from the standpoint of what women need to know, the best models of screening programs, manpower, risk stratification, basic research, and payment. These contributions could then be documented and distributed widely and could continue to draw attention to and expand the reach of the report and the salience of breast cancer screening and early detection and the development and deployment of new screening technologies.

[1] The Apex Foundation support was given in memory of Mabel Frost McCaw and Joan Morgan, and in honor of Sallie Nichols, Beth Weibling, Jane Carson Williams, Bonnie Main, and Amy McGraw.

The one day symposium reported here was designed by a planning group that relied on the experience and wisdom of the staff of the Breast Cancer Screening Consortium at NCI, the American Cancer Society's Cancer Screening staff, the CDC, the medical leadership of The Breast Cancer Research Foundation, and members of the Board. The morning of the symposium featured an overview plenary session introduced by Evelyn Lauder, Founder and Chairman of The BCRF; and Dr. Edward Penhoet, Chairman of the Committee, with presentations from senior experts in cancer research and care, breast cancer screening and epidemiology, radiology and breast imaging, basic cancer research, and national payment policy.

The afternoon consisted of invited panel and group discussions on two major topics important to breast cancer detection, delivery of screening services and technology development and deployment. A brief wrap-up session at the end of the day allowed two rapporteurs of the group discussions to summarize the information and recommendations presented during those sessions. They were followed by a final commentary from a representative of cancer survivors. The agenda identifying the morning speakers, their titles, affiliations, and topics, and the afternoon panelists with their titles and affiliations, can be found in the appendix. The speakers in each group discussion were assembled from different governmental, academic, and private sector organizations to provide a wide range of perspectives. The participants in discussions, questions, and answers during the morning and afternoon are also reported.

Following this Introduction, Chapter 2 opens with introductions from Mrs. Lauder and Dr. Penhoet and presents the remarks of the morning plenary speakers in order of appearance along with the question and answer sessions, Chapter 3 presents the speakers and discussions from the afternoon, and Chapter 4 concludes the symposium with summing up from the rapporteurs and a survivor representative and some further discussion. All the presentations and discussions were edited for easier reading and to add graphic material in the form of figures numbered sequentially from PowerPoints used during each speaker's presentation. This dissemination report contains only what was said and displayed at the symposium. It is, therefore, a less formal forum than a Board or IOM report. Much interesting information and analysis and provocative ideas and suggestions can emerge during such an event from the experts, officials, and opinion leaders assembled. The Board and The Breast Cancer Research Foundation hope that this record of the day will provide continuing food for thought and ideas for actions in support of breast cancer detection and diagnosis and saving women's lives.

Roger Herdman
and
Larry Norton

2

Plenary Session

Introduction to the Symposium and of the Founder and Chairman of The Breast Cancer Research Foundation

Edward Penhoet, Ph.D.,
Chair, Committee on Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis; and Director, Science and Higher Education Programs, Gordon and Betty Moore Foundation

Good morning and welcome to this symposium to discuss the product of almost two years' work on saving women's lives. We are delighted to see so many who have joined us. Before we begin the formal presentations, however, we have a special guest, Evelyn H. Lauder, who has come from New York to say a few words to us. Mrs. Lauder is the Founder and Chairman of The Breast Cancer Research Foundation which has supported work at the Institute of Medicine's (IOM's) and National Research Council's (NRC's) National Cancer Policy Board on breast cancer research for several years, and in particular has been an important supporter of this project and co-sponsor of this symposium.

Introductory Remarks
Evelyn H. Lauder
Founder and Chairman,
The Breast Cancer Research Foundation

Thank you, Dr. Penhoet, and good morning, everyone. I am very flattered to be introducing this symposium and pleased, on behalf of The Breast Cancer Research Foundation to welcome all of you. We are all here because we share a common goal, to save women's lives from breast cancer. As founder and chairman of an organization that has raised over $95 million since 1993 to support innovative research in preventing and curing cancer, I know and appreciate the critical role of early detection and diagnosis.

In 1992, along with Alexandra Penney, who was then editor of *Self* magazine, we introduced to people all over the world the pink ribbon which has come to be recognized as a universal sign of breast health and awareness. I'm proud to say that since that time, the Estee Lauder Companies alone have distributed over 45 million of these ribbons at our counters worldwide.

In 1993, I led a delegation of Estee Lauder executives and editors from *Self* magazine to Washington, D.C., and we raised a window shade on which had been pinned 250,000 names. Through coverage of this event in the national press and television and our visit to Hillary Clinton at the White House, we drew attention to the fact that the federal government needed desperately to give more funds for breast cancer research. It was then that Major General Travis was designated to head a study as a result of which substantial new funds were made available for breast cancer research. So from the outset, we have been dedicated to supporting clinical and genetic research into the causes and treatment of breast cancer.

Our Foundation's grants for stellar research projects have really grown. For example, and of particular interest today, we provided major funding for the 2001 IOM and NRC report, *Mammography and Beyond* (Institute of Medicine and National Research Council, 2001). After its publication, Dr. Larry Norton, our scientific diretor, who is with us today, addressed the Foundation's board and told us that the IOM was appointing a new committee to embark on a study to expand on that report. Dr. Herdman and Dr. Joy can attest to the enthusiasm with which Dr. Norton's suggestions were greeted. We called the IOM right after the meeting to say that $100,000 had been pledged on the spot by members of our board for the new committee's work.

Since then, the Foundation has provided steady financial support for the project and has eagerly awaited the committee's findings, which were released to the public last week, and are being expanded and presented in greater detail at this symposium today.

I could not be more proud of the research that Dr. Norton and his colleagues on our Medical Advisory Board have recommended for support by The Breast

Cancer Research Foundation. In fact, the first scientific presenter this morning will be our dear friend, Dr. Laura Esserman, whose research we have also supported for 10 years.

The work that you have all done in assessing the state of breast cancer and early detection in this country and in identifying ways to improve detection and diagnosis is of major importance. Your research will make a huge difference in women's lives, and for that, I want to thank you personally. You have fueled the determination of volunteers like myself and Peg Mastrianni, the deputy director of the Foundation, and her colleague, Anna DeLuca, who directs public affairs. You encourage us to work toward increasing public awareness and support, though magazine editorials, newspaper reporters, as well as fundraising. So I can't thank you enough.

DR. PENHOET: Thank you, Mrs. Lauder, for your comments and especially for the opportunity you have provided all of us on this committee to work on this project, to join you in the fight against breast cancer. We would also like to acknowledge the other sponsors of our work and the symposium: the Broad Reach Foundation, the Apex Foundation,[1] and the National Cancer Institute.

In addition, I would like to extend our warm thanks to Dr. Janet Joy, who was our report's study director. It is hard to imagine anybody working harder for a period of 18 months; she has produced a very fine, readable report. I also recognize the vice chair of this committee, Dr. Diana Petitti, who worked closely with me and with Janet throughout the entire process. Diana, thank you so much for your help with this work today. Due to the shortness of time, I won't go through the entire roster of participants in this committee, who are listed in the front of the report. This was an extraordinary group of experts who worked really hard to achieve the tasks assigned to this committee. We are grateful to all the participants, especially the sponsors and the staff of the IOM, for bringing us to this point.

The report that we are here to discuss is the result of an 18-month study charged with examining existing and evolving approaches—that doesn't mean just technology—that hold the greatest promise for improving the early detection and diagnosis of breast cancer. The committee focused on identifying which approaches are likely to save the most lives in the near term. This includes technology in the broadest sense—from specific tools such as digital mammography, MRI, biomarkers, and proteomics to how these tools and strategies can be most efficiently deployed in clinical practice. The committee's recommendations address what we thought were the most important steps that could be taken to improve outcomes of breast cancer in the near term.

First, we have not yet optimally used the most powerful tool at our disposal, that is, mammography. So a number of our recommendations relate to the im-

[1] The Apex Foundation support was given in memory of Mabel Frost McCaw and Joan Morgan, and in honor of Sallie Nichols, Beth Weibling, Jane Carson Williams, Bonnie Main, and Amy McGraw.

provement of the practice of screen-film mammography and to better access to mammography.

Second, the committee believes we need technology and procedures to develop individually tailored screening strategies so that high, medium and low risk individuals can receive the type of screening that is most appropriate to them. This poses a difficult task—to stratify risk in the population as a whole. We think that the promise of genomic technology has already been realized in a few instances in breast cancer, for example, the BRCA family of genes, and that in the future this might be expanded significantly. Our ultimate objective would be to customize and optimize screening strategies for individual women.

Third, we need to address the weakest link in the pathway of technology development, that is, demonstration that a new technology or procedure truly improves health outcomes. Here, we recommend the formation of centers in the United States, either real or virtual, to integrate new technologies, particularly to integrate the basic research findings in biomarkers and proteomics, among other advances that we discuss, with clinical practice, and then once those things have become integrated, to make sure that clinical utility is demonstrated in a convincing way.

We believe these recommendations are fully consistent with the new initiative at the National Institutes of Health (NIH), the road map, which seeks to better integrate basic research with clinical practice.

The purpose of this symposium then is to discuss the implications of the recommendations in this report, as well as how they might be implemented. Several members of the committee are here today. You will hear from some of them, as well as other experts who have not been directly involved in the report. This symposium also will provide an opportunity to discuss the issues and complexities surrounding the early detection of breast cancer in much greater depth than is possible in a press conference. This morning we will hear from a series of speakers who will be addressing different themes following the outline in the report. In the afternoon, there will be two concurrent group discussions, ending with a plenary wrap-up discussion.

The Pros and Cons of Screening Mammography: What Women Need to Know—An Overview of the Report's Findings on Mammography

Laura Esserman, M.D., M.B.A.,
Director, Carol Franc Buck Breast Care Center
and Professor of Surgery and Radiology
University of California, San Francisco

The first thing that women need to know is that mammography is early detection, not prevention. The World Health Organization has set out principles for

detection through population-based screening. The disease should be serious and prevalent, like breast cancer. There should be a detection test that is sensitive and specific, well tolerated, inexpensive, and that changes therapy or outcome. This is important, because population-based screening is very different from individual screening. We are going to talk about what that means, and also how our new understanding of the biology of breast cancer should affect our approach to screening. The goal for breast cancer screening is not to detect every possible breast disease or abnormality but rather to prevent deaths from breast cancer. That is the ultimate test of whether screening is of value.

What are the pros and cons of mammography? There have now been seven randomized trials that demonstrate 20 to 30 percent reductions in the relative risk of breast cancer mortality depending on age at screening. Mammography finds cancers at an earlier stage than detectable by physical examination. Small cancers are less likely to metastasize and are therefore more likely to have good outcomes. Small cancers are also more amenable to breast conserving treatment approaches and better cosmetic results. There really is no other technology that has been shown to systematically find tumors at an earlier stage, and 12 countries have implemented systematic screening programs.

Participants in the global mammography summit in June 2002 reviewed the available evidence and unanimously decided that there was no reason to change screening programs. But they also noted that mammography is only one part of the total management of a woman with breast cancer; integration with further diagnosis and treatment is critical.

So, why should there be any controversy? Well, mammography doesn't find all cancers. The sensitivity is 83 to 97 percent; it depends on age and probably breast density as well, which is related to age. Mammography has a relatively high false positive rate, which is important. Ten percent is high for population-based screening. Mammography is resource intensive. Quality is actually quite variable, depending on how mammography is performed.

The absolute number of women benefited by mammography is very different from the relative risk reduction. There is perhaps a four to six percent reduction in absolute mortality as opposed to the 20 to 30 percent reduction in relative risk. The absolute reduction also depends on age, and the value is quite a bit lower for young women, perhaps in the 1 to 2 percent range.

Finally and importantly, finding cancers early does not guarantee cure. Biology can trump detection, and better understanding biology is an important theme of this report. Mammography has also led to a very large increase in the detection of in situ cancers; some call this the friendly fire in the war against cancer, an apt description, I think.

Do we just need a more sensitive screening detection tool? Maybe we should be using ultrasound or magnetic resonance imaging (MRI), even though they are three- to fifteen-fold, respectively, more expensive than screen-film mammography. MRI, for which there has been considerable enthusiasm, can identify tumors that don't form masses and tumors in dense breast tissue. MRI certainly is useful when we know someone is at extremely high risk for developing breast cancer. For women with genetic susceptibility, some of whom are known to have an 80 percent lifetime risk of developing breast cancer, MRI, even though expensive, has been shown to be much more sensitive than screen-film mammography, particularly because those being screened are young women with dense breast tissue in whom underlying breast tumors are more likely to occur (Kriege et al., 2004)

The problem is that MRI is too sensitive. It finds all kinds of things that aren't cancer, but whose significance is unclear, and its performance in detecting cancer is not better in women with fatty breast tissue who make up the majority of women ages 50 to 70 for whom screening is designed. Furthermore, biopsy tools are not readily available. Lastly, it is also too expensive for a general population screening test. That is probably also true for ultrasound, which is very labor intensive.

These factors bear on how we want to think about population-based screening with the objective of saving women's lives. Screening has enormous impact. Economically, mammography in aggregate U.S. cost is somewhere in the six to ten billion dollar range. Emotionally, women who are called back for low risk lesions or women with indolent disease, who assume they have life threatening disease, pay an enormous price, if in fact these things would not have otherwise come to clinical attention.

The sensitivity and specificity of mammography depend to some extent on who interprets the image. High sensitivity, that is, the chance that a mammogram of a woman with breast cancer will be correctly interpreted as positive for cancer, is clearly desirable. But if you find absolutely everything, at some point you are going to pay the price of a high false positive fraction, or 1 minus the specificity (the specificity being the chance that the mammogram of a woman without breast cancer will be correctly interpreted as negative for cancer).

Historically, it was thought that this was just a simple tradeoff, but it is not. Sensitivity and specificity are highly variable. Some breast imagers are more experienced and/or skilled than others, and, as exemplified in Figure 2.1, these imagers (represented by the curve for "better" interpreters) have better ratios

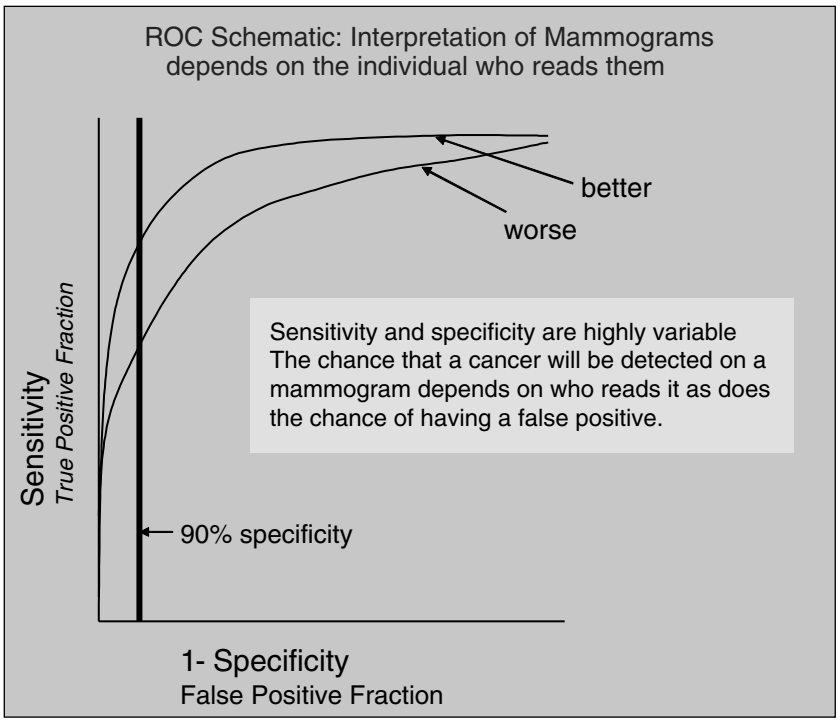

FIGURE 2.1 Better interpretation means a better ratio between true and false positive cancer readings as shown in this ROC (receiver operating characteristic) schematic.

between true positive and false positive readings of mammograms. In short, the chance that a breast cancer will be detected depends in part on who reads the mammogram, and the chance of having a false positive also depends on the quality of the interpretation. (Of course, other factors are important, as well, such as the quality of the image, positioning and compression of the breast, among others.)

In the U.S., over 75 percent of biopsies following mammography are not cancer. Although, there is variation in biopsy rates internationally, variation may be greater in the U.S. perhaps than in countries where they have more focused screening programs.

The organization of care clearly affects the quality of screening. High volume, experienced mammographers find the most cancers and miss the fewest

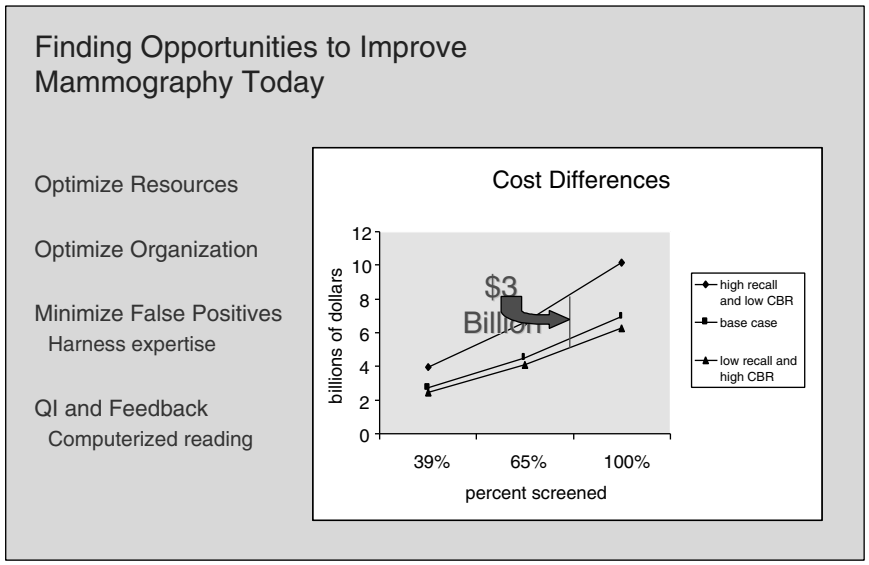

FIGURE 2.2 Cost differences between programs of varying efficiency. CBR = cancer to biopsy ratio.

cancers. Furthermore, coordinated teams make sure that, in the trajectory of care, the right procedure is done at the right time. We think that integration of the various aspects of care and feedback (learning from experience) is critical to optimizing performance.

Where there are focused, organized screening programs, the fraction of positive operative breast biopsies can be 80 to 90 percent (UK) or 85 to 95 percent (Sweden) compared to the U.S. fraction of 20 to 70 percent. Data such as these suggest the value of high volume screening programs and support the conclusion that in some countries, mammographic interpretation is more consistent and of greater specificity than in the U.S. Such effective, highly-organized programs can also be more efficient with the potential for significant cost savings as indicated in Figure 2.2, which shows the high aggregate cost differences between high recall, low cancer to biopsy ratio (CBR) and low recall, high CBR programs. Higher quality can also be much more cost-effective (Burnside et al., 2001).

We are also beginning to understand, through molecular fingerprinting, that all breast cancers are not the same. We can tell the types of cells from which breast cancers arise—where they are in the milk duct—and these different

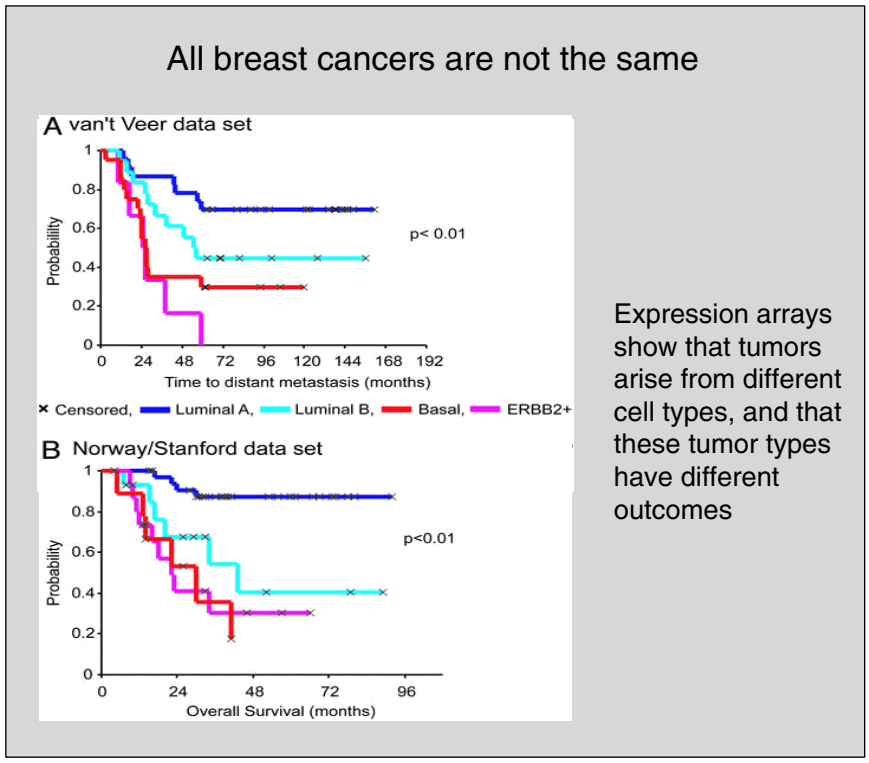

FIGURE 2.3 All breast cancers are not the same.

sources affect the outcomes for patients. As the survival curves in Figure 2.3 show, some types of breast cancer have a much higher risk of progression to distant metastasis and death (Sorlie et al., 2003).

For example, new data from the Women's Health Initiative, shown in Figure 2.4, indicate that some populations, such as African American women, have different types of breast cancer, so that, although the frequency may be lower, more aggressive, poorly differentiated plus estrogen receptor negative tumors are much more frequent.

Breast cancer from milk duct luminal cells is more frequent in older women and is more likely to be well differentiated and amenable to treatment. Breast cancer from basal cells, which is more frequent in younger women, is often more aggressive and less likely to be amenable to treatment. Patient populations which are more likely to have aggressive, fast-growing tumors that are discovered when they are larger or have spread and are, therefore, less responsive to treatment, are not as likely to benefit from screening. That is relevant to what Dr. Penhoet was saying about screening and stratifying risk.

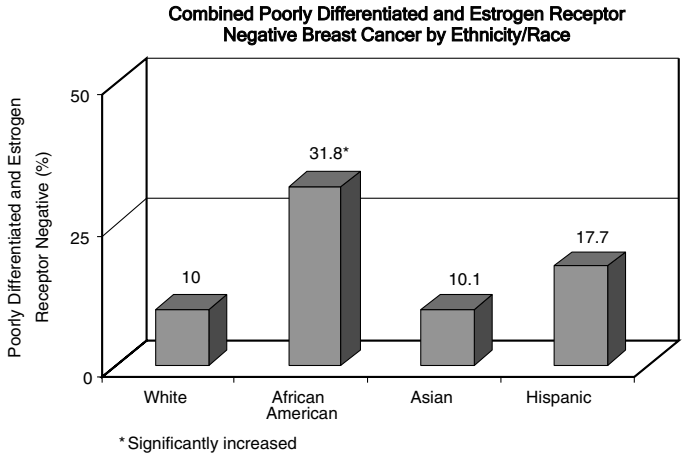

FIGURE 2.4 African American women have significantly more aggressive, poorly differentiated, estrogen receptor (ER) negative breast cancer.

So the lessons from biology are that tumors grow at different rates. Fast growing tumors may not be caught in time by mammography. Other slow growing tumors provide time to do interval screening and detection before they are a threat to the patient. And there are still other cancers that may grow so slowly, particularly if they are in older women, that they may never come to clinical attention. So part of our strategy has to be designed in concert with our new understanding of the biology of breast cancer.

Returning to the implications of less frequent but more aggressive breast cancer originating in basal cells in African American women, one would expect perhaps that screening would make more of a difference in Caucasian women—more cases to find, more chance of making a difference through early detection. This does appear to be true, at least in part, as indicated by Figure 2.5 from the IOM report. Breast cancer incidence in African American women remains lower than in Caucasian women over time, but mortality is greater and has not been as affected by increasing rates of screening. While other factors may explain some of these effects, such as access and treatment variables, it is also important for us to integrate our understanding of biology when interpreting data like these. We should not expect too much from mammography. We know that mammography is not going to make a difference for all breast cancers, and breast cancer is not one disease. And mammography cannot be expected to find all cancers either.

The cost of mammography often exceeds reimbursement, and this is also important in thinking about strategies. Having run a digital mobile van service

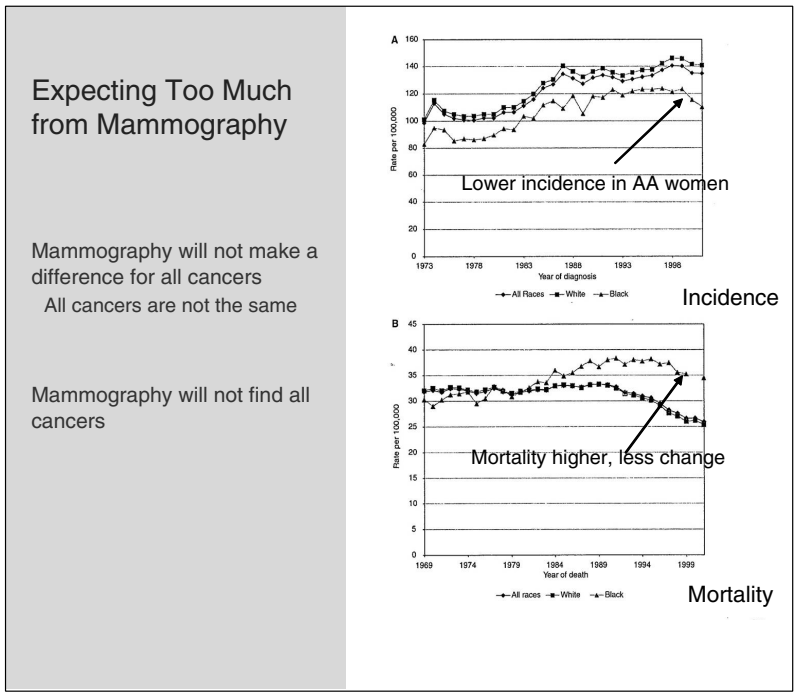

FIGURE 2.5 Consider biology when interpreting mammography data.
SOURCE: Surveillance Epidemiology and End Results Program, NCI.

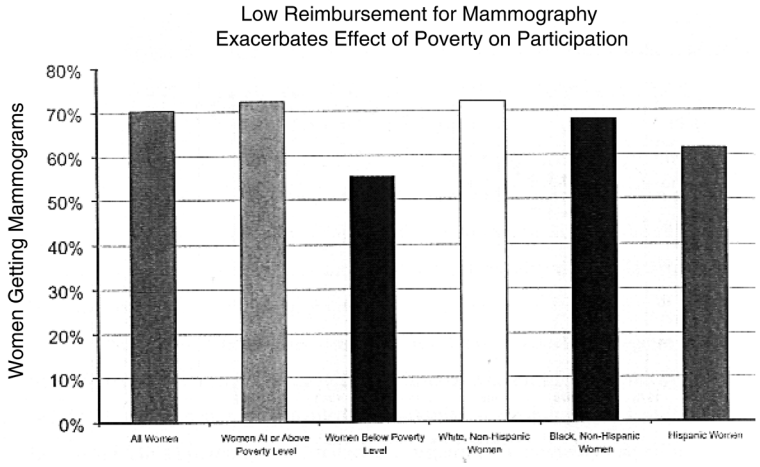

FIGURE 2.6 Poverty is the greatest barrier to mammography screening.

for underserved women, I know just how significant reimbursement is. If you want to provide mammography for the underserved, you will need to raise a lot of money. Figure 2.6 shows that poverty, not race, is the greatest barrier to screening.

Age is another important consideration in screening strategy. Mammography is most effective in women 50 to 70 years of age. Returning to the biology, we know that this is a population of women who are at greater risk for breast cancer and have cancers with slower growth rates. Unfortunately, screening rates decline in older populations. Less than two-thirds of women 65 to 70 are getting screened. In late age perhaps other competing risks of death make breast cancer less important. But the cost-effectiveness of screening mammography is at least three to five times higher in women aged 50 to 70 compared to those aged 40 to 50. In the younger women, the risk of having cancer is lower, the sensitivity of mammography is lower, the recall and biopsy rates are higher, and there are probably more of these tumors that metastasize early. Some of this information is summarized in Figure 2.7.

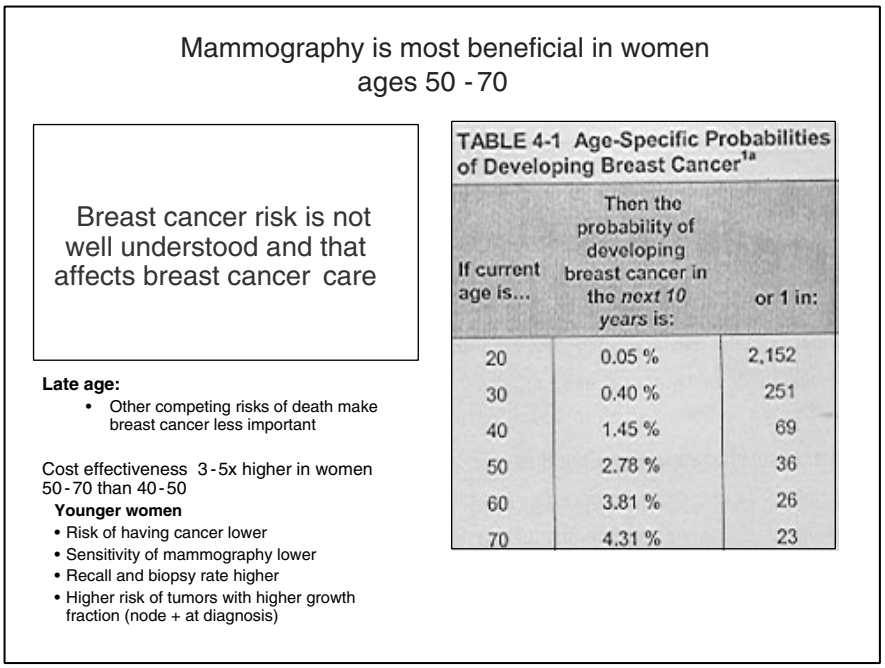

FIGURE 2.7 Advancing age and mammography.
SOURCE: Breast Cancer Facts and Figures, updated to 2003-2004, American Cancer Society.

What do we do with the information from mammography? The image interpretations are reported as BI-RADS® scores (Breast Imaging Reporting and Data System™), which is the standard way of classifying mammograms. A BI-RADS® 5 means probably (>75 percent) cancer, and a BI-RADS® 3 means probably benign (1-2 percent chance of cancer, 6 month follow-up recommended). A BI-RADS® 4 is called suspicious, and women are sent a letter to that effect. But there is enormous variation in the odds of having cancer in this category—from 3 to 75 percent. It might be an in situ pre-cancer, or it might be an invasive cancer, and it might often lead to biopsy and contribute to the false positive rate and create quite a bit of alarm. Some of this could be avoided by framing the information properly and having some self control. A new classification is going to have BI-RADS® 4A, 4B, and 4C, which is going to stratify the suspicious category, but it will take some time to disseminate a new system and determine what effect it will have.

My friend Gilbert Welch in his book, *Should I Be Tested for Cancer,* which I strongly recommend to anyone who is interested in this topic, points out that in our zeal to make a difference through screening, we need to be sure that we are making the right difference (Welch, 2004). Screening is a good thing if the incidence of detected late stage cancers is going down, and the incidence of detected early stage cancers is going up, even if the total incidence of cancers stays the same. But a stable incidence of late stage cancers with an increasing incidence of early stage cancers may indicate a problem. We may be seeing a misleadingly higher survival rate which is only due to inclusion of more early stage cancers, reflecting an increase in ductal carcinoma in situ (DCIS), that is, tumors that might not have had any consequences for the patient. Actually, presumably as a result of screening detection, the incidence of invasive breast cancer has increased and of late stage cancer decreased somewhat, but the incidence of DCIS has gone up dramatically. That is why biology becomes such an important part of our recommendations—to think about what it is that we are detecting, and how to treat it, making sure that we use the screening information judiciously.

Ductal carcinoma in situ is a disease in which the cells lining the lumen of the milk duct look like cancer, but they have not developed the capacity to invade. As such, they are 99 percent curable, but they comprise a lot of the breast cancer cases that are detected through mammography, almost 50,000 women a year. These are healthy women at high risk for invasive cancer in the future who don't have life threatening disease at the moment. We need to be careful about what we do with this information.

The mammography controversy was partly fueled by the fact that finding more DCIS was generating increasing mastectomy rates. These have subsequently decreased because more patients are being treated with breast conservation. But I feel that we have an incredible opportunity to use DCIS, now that we have this biomarker, to figure out how to prevent breast cancer if we would have

patience and stop operating right away on these patients. As a surgeon, I feel I can lead the charge in this area. As technology finds evidence of cancer earlier and earlier, we should accompany that by parallel treatment and prevention strategies that are appropriate.

What are the solutions that the IOM report recommends? The first is that if you are going to screen in the first place, then screen well. Screening mammography can definitely be improved. The chance of being called back for an abnormal mammogram is as high as one in ten. The chance that a woman will have a breast biopsy or some kind of workup for an abnormal screen after 10 years of screening is one in two. Breast cancer at any given screening is a low frequency event, 1 or 2 cancers per 1,000 mammograms from women ages 40 to 50, and 5 to 7 cancers from women ages 50 to 70. And the quality of mammography is variable.

At the same time, our ability to screen well is threatened. The number of radiologists willing to read mammograms is declining. Well trained mammographers are in short supply. The cost of mammography exceeds reimbursement. Even though we know that well-trained mammographers find more cancers and have fewer false positives, probably less than a third of our mammograms are being read by such individuals.

Some of our solutions involve focusing on leveraging emerging technologies, better organization of services, and quality improvement. Risk stratification of women might reduce the volume of mammograms and allow us to focus on women with the greatest frequency of cancer. Mammographic services and interpretation concentrated in centers of excellence that are integrated into multidisciplinary care are an important part of the solution. New technologies such as digital mammography create opportunities for change such as the addition of technologies like computerized decision aids which can reduce the difference between the average mammographer and the expert mammographer.

We recommend considering adoption in the United States of elements of successful breast cancer screening programs from other countries, including centralized expert interpretation, regionalization of programs, outcomes analysis, and benchmarking. We think it is important to collaborate with health-care providers and payors to improve quality; to develop and adopt practices that promote self improvement; to develop and disseminate technologies, such as computer aided diagnosis that will improve quality; and to expand the capacity of breast imaging specialists by specially trained allied health personnel.

We recommend integrating biology, technology, and risk models to develop new screening strategies for breast cancer. This will involve harnessing emerging molecular applications to not just find cancer, but to determine what type is present or likely to develop, and enable reasonable predictions about treatment and outcomes. Some of these applications may emerge from proteomics. Examination of patterns of proteins in the blood as a way of screening raises exciting possibilities that will require careful development and application.

The two major drivers of cost in mammography are the percent of women screened and the cost per mammogram. A cheap easy test that would allow us to limit mammography only to those women considered to be at risk would be a home run. So it challenges us to think not just of having competing technologies for mammography, but to integrate them in a better way. In the future, tests might be layered for optimal effect: screening by proteomics; susceptibility testing; inherited genetic variations (BRCA mutations, SNPs); or nipple fluid aspirate analysis to identify the risk of having or getting breast cancer, then using imaging techniques, like mammography, MRI, or ultrasound, as localizing tools or secondary screens; and finally using some of these techniques (MRI, PET, various probes and biomarkers) further to monitor response to therapeutic interventions. Harnessing risk discriminators and biomarkers together in a multidisciplinary way would give us the power to make progress in guiding screening and intervention strategies. Expression profiles, looking at circulating tumor cells, and characterizing these tumors might help us understand their prognosis and how they might respond to therapies or help us generate new ideas for targeting therapies.

Of course, to move in this direction, we will also need to ensure that these concepts of risk are better taught. Decision aids are going to be needed. We have recommended that research funders help develop tools that facilitate communication regarding breast cancer risk to the public and health care providers, so that we really understand the various risks and benefits, including the risks associated with emerging biomarkers. That means finding ways to teach women about their risks and the benefits of interventions. All development, of course, should be in the context of what is currently clinically practiced, because otherwise, you may find that what you are developing is not terribly useful.

Our final recommendations involve improving the environment for research and development of new technologies for breast cancer detection. In the report, we have highlighted the need to try and create centers where people from the research, imaging, and screening communities are working as a team to put all the pieces together. We know that the optimal use of new technology requires specific attention to implementation and dissemination. We were fortunate to have had members of our committee that were very much involved with operations management, or the dissemination of technology, and we understood that perhaps the hardest thing of all is changing practice. So paying attention to how we re-engineer care, help people redefine their jobs, and how we monitor that change, is absolutely essential.

In conclusion, the goal of early detection needs to be integrated with subsequent strategies. Understanding risk, applying early methods to find out if a cancer is present, understanding the likelihood of disease progression or whether someone can be left alone, understanding what tests can tell about a woman's responsiveness to systemic therapies or whether new strategies are needed, de-

veloping the idea of targeted prevention and therapy, we believe these are among the strategies that will lead to saving women's lives.

Challenges to Expanding Mammography: Better Quality for Women in Screening Sites

N. Reed Dunnick, M.D.,
Professor and Chair, Department of Radiology, University of Michigan

Today, I want to talk to you about mammography, both screening and diagnostic, and ultrasound. Radiologists also perform a variety of procedures—aspirations, core biopsies, needle localizations, and now getting into ablation techniques. In essence, the breast radiologist is the primary care physician for breast disease.

The challenges before us regarding access to mammography include a human resources shortage, high liability risk associated with mammography, and relatively low reimbursement. The countermeasures include producing more radiologists, improving their productivity (work harder, physician extenders, computer assisted diagnosis), initiating tort reform, and adjusting the payment schedule.

Let's start with the human resources shortage. Demand for radiology services is rising at about one percent per year because of population growth. In addition, demographic changes, like aging of the population, may increase the need for radiology services by a further half a percent a year.

The biggest change however has been in medical practice. The sophistication of imaging technologies has reduced the value of the clinical history, has made the physical examination almost trivial, and has provided a tremendous amount of essential data for diagnosis and treatment. We have seen a shift from plain-film radiography to cross-sectional imaging techniques. Although these techniques are more expensive, in terms of information per dollar spent they are actually less expensive than the older techniques.

I summarize in Box 2.1 that there is approximately a 6 percent increase in growth of radiology services annually. Unfortunately, we are not producing radiologists fast enough to meet this demand. We train approximately a thousand new radiologists each year, but about 500 retire each year, so we have a net annual increase of only approximately 500 radiologists. If we assume that there are 33,000 practicing radiologists, then the growth rate in radiologists is only 1.5 percent per year. Demand up 6 percent, supply up 1.5 percent—we have a gap of 4.5 percent each year.

BOX 2.1
Demand for Radiology Services Is Rising

	Percent
Population growth	1.0
Evolving demographics	.5
Change in medical practice	3.0
Exam annual growth	4.5
Shift to CT, US, and MRI	1.5
Work annual growth	6.0

How did we get into this situation? In the early 1990s, with the spread of managed care, we began to hear predictions that the need for ancillary services in health care systems would decline by 30 to 50 percent. At the University of Michigan, one analysis (Billi et al., 1995) predicted that the 45.2 full-time-equivalent (FTE) faculty that we had in our department in 1992 would drop to somewhere between 12 and 16 FTEs as a result of this reduced utilization of ancillary services secondary to managed care. That prediction was a bit off. We are now at 80.4 FTEs and climbing.

So what can we do? The first thing would be to train more radiologists. The problem here is that the Centers for Medicare and Medicaid Services (CMS) capped the number of positions allowed for graduate medical education payments under Medicare at their December 1996 levels. As a result of the managed care scare, many programs reduced the number of radiologist trainees. Some reductions were voluntary, in anticipation that there would be insufficient radiology jobs, and others were less than voluntary as the institutions converted some radiology positions to primary care positions.

A number of programs were actually eliminated. They felt they were unneeded. They may not have been doing a very good job in the first place, and maintaining the residency training programs was not really worth the effort. Most of these were not in our university programs, but in private practice or multi-specialty clinic settings.

So this cap that CMS imposed froze those reimbursable positions at radiology's nadir, and now we are stuck. We can no longer respond as a normal market economy; we are limited by that artificial cap. Box 2.2 shows the number of residency positions offered by the radiology matching program, the precipitous decline in the late 1990s, and the gradual slow increase thereafter. We have increased the number of residents at the University of Michigan twice, and a number of other programs have also been able to do that, but we are still not back to the levels we had prior to this managed care scare. Over this time period, the

BOX 2.2
Residency Positions Offered

Year	Positions Offered
1996	1,154
1997	890
1998	843
1999	852
2000	841
2001	875
2002	920
2003	979
2004	981

NOTE: Residency positions in radiology offered through the matching program fell dramatically after 1996 and have not recovered.
SOURCE: National Residency Matching Program.

percent of the radiology positions filled through the matching program has risen to 99 percent, and the final percent or two fills from the pool of applicants who did not match, so all residencies are filled.

International graduates comprise a second potential source of radiologists, but there are difficulties in identifying and assessing the credentials of candidates. They train at institutions and in systems with which we often are not familiar, and it is not easy to determine how good they may be or to interview them. Once identified, there is the increasingly difficult visa problem and the different state medical licensing rules and procedures. And finally, board certification is yet another hurdle that is a particular problem for mammography because of Mammography Quality Standards Act (MQSA) regulations.

If we cannot increase the size of training programs or recruit international graduates, could we retard the retirement of existing radiologists? There are possibilities here. Some radiologists may be ready to retire, but perfectly happy to continue in a part-time rather than a full-time capacity. We can also allow subspecialization. Many people like to do one thing, but don't like to do another thing. We can make their job description more focused and induce them to stay within the work force.

It is worrisome that the human resource shortage is affecting academic programs more than private practice, and that it affects mammography more than other areas in radiology. These are discouraging trends, because academic medical centers are training the next generation, and if we agree with Dr. Esserman

that it would preferable to have mammography services delivered by experienced, high-volume breast imagers, the academic centers are the places with that expertise and volume.

Mammography is affected disproportionately by the trends mentioned. The problem is that radiologists choose other fields. If there are not enough radiologists to do all the work available, which work would you choose to do? Would you choose a field in which professional liability is high, reimbursement is low, and regulation is significant, or would you choose something else? Radiology help-wanted ads reflect this. Over the seven year period from 1994 to 2001, the fraction of these ads specifically looking for mammographers has grown disproportionately, from 6.4 to 10.2 percent. This happened during a time when want ads for radiologists of all kinds were more than doubling, so the actual numbers of ads for mammographers grew dramatically (Saketkchoo et al., 2002).

The malpractice issue is very significant for radiologists but particularly for mammographers. I wish the data on this were more up-to-date; studies are primarily from the 1990s. They show that misdiagnosis is one of the most common causes of malpractice litigation in radiology. Although bone disease, typically fractures, was the most common at the time of the study (Berlin and Berlin, 1995), breast cancer was a close second and was rising at a rate that suggests that if data were more recent than 1994, breast cancer would surely be number one in terms of liability.

In discussing mammography and reimbursement, we should be clear about the differences between screening and diagnostic mammography. Screening means the patient is asymptomatic; the exam is performed by a technologist, and the radiologist reads this at a later time. This allows him or her to read a large batch of screening mammograms in a quiet, undisturbed environment, and that can be relatively efficient. A diagnostic mammogram, on the other hand, means that the patient has a history of breast cancer, or an abnormal physical examination—a mass, bloody discharge, pain—or a prior abnormal mammogram. This is a real-time process. The radiologists must be present to look at the films. They may ask for repeat or different views, and then will then look at those, too. There is constant interruption in this process, and it is relatively inefficient.

This difference is not really reflected in either the relative value unit assignments to these procedures or the resulting reimbursement. Although Medicare fees vary somewhat in different regions, we can use Michigan as an example. For our carrier, the professional component of the fee for screening mammography is $40 and for diagnostic mammography $49. Our mammography group would say that the difference in effort between these two is three to one, not five to four. So we think there is an imbalance here.

Furthermore, compare this with other radiology options, for example, reading an abdominal CT scan, for which the payment is $72, or reading a head MRI

for which the professional component is $134. Clearly, mammography is not economically attractive.

The next thing we could do to enhance access is increase the radiologists' productivity. Radiologists could work harder. We could use physician extenders. We could take advantage of technology, and we have computer assisted diagnosis systems that might be employed.

In recent times, radiologists' work loads have been increasing (Bhargavan and Sunshine, 2002), most notably in academic radiology (23 percent), but also in multi-specialty practice (17 percent) and even private practice (8 percent). Furthermore, a recent survey (Sunshine et al., 2002) reports that a majority (51 percent) of radiologists believe they already have too much work. These findings do not encourage us to think there is a likelihood that productivity will close the gap between supply and demand.

What kind of physician extenders do we have that might address the access problem? I think that we should have each job done by the person specifically trained to do that job. So I want radiologists spending as close to 100 percent of their time as possible doing work that only radiologists can do. I do not want them hanging films, retrieving old films, filling out quality assurance forms, or looking up data. Ultrasound in our institution is done by the radiologist, whereas ultrasound generally is done by a technologist. We need to train more technologists to do breast ultrasound, rather than having the radiologists do that. And physician extenders could take on the hugely important task of patient education.

What else might these physician extenders do? Could they prescreen screening examinations? Might they identify all of the clearly normal mammograms enabling the radiologist to move even more quickly through them for a final interpretation? Might they flag abnormal examinations, and, by adding another set of eyes, reduce the number of misses by the radiologist? Would this in essence be double reading? There are a number of publications that report that technologists and other non-physicians can be taught to do that (Hillman et al., 1987; Sumkin et al., 2003).

There are technologies that we could employ to make us more efficient. PACS (picture archiving and communication systems), which allow the storage and transmission of images, have great potential. This requires that the radiology be digital, and we are anxiously awaiting the results of the trial that Dr. Pisano is leading through the American College of Radiology Imaging Network comparing digital and conventional film-screen mammography in almost 50,000 women.

Voice recognition allows you to dictate your report at the time you remember the examination and have it ready right away; there are a number of ways that we could make that more efficient. Electronic medical records allow the use of a computer at the radiologist's side to look up any pertinent information on the patient. In fact, I think we can even do better than that. Relevant literature

could be reviewed at the same time as a particular case is seen, including what patients with the same diagnosis look like. Electronic teaching files make it possible, while examining a particular image that might be amyloid of the breast, to review a series of patients with breast amyloid to see if the features are the same.

Computer assisted diagnosis (CAD) is something that many of us are quite optimistic about. These systems mark suspicious areas like micro-calcifications or masses. In one study (Astley et al., 2002) CAD was able to detect 87 of 90 cancers which is excellent. On the other hand, it also marked 556 of 810 normals. Would this help improve the sensitivity? My guess is that it might. On the other hand, you still need a radiologist to see what your CAD system is doing, so that you don't call back two-thirds of your patients.

In summary, my recommendations are as follows. To address the problem of too few radiologists, raise the CMS cap on house officer positions. This could either be an institutional increase, in which case radiology would be competing with all of the other subspecialties because the demand for more physicians in many other areas has not decreased since 1996, or it could be specifically targeted to radiology.

Second, to encourage radiologists to provide mammography, the problems of high liability and low reimbursement need to be addressed. We need tort reform and an adjustment of the reimbursement schedule.

Third, to increase the productivity of the radiologists we have, we should continue technology development. There are a number of exciting possibilities. I am optimistic about digital mammography. Our electronic information systems have been a great help in many areas in radiology and of course, I am looking forward to the use of CAD more extensively throughout the country. I think these technologies may help all of us save women's lives.

Better Models for U.S. Mammography Services: Implications for Accuracy and Encouragement of Screening. Better Quality Through Better Organized Mammography

Robert Smith, Ph.D.,
Director of Cancer Screening,
American Cancer Society

What might we achieve, or not achieve, through better organized screening? Much of this presentation grew out of a project supported by NCI, CDC, and the American Cancer Society and led by Dr. Helen Meissner at NCI, that we have been working on for the past two years called Lessons Learned About Cancer Screening, published as a supplement in the journal, Cancer (Meissner et al., 2004).

Screening is not a single event, but rather a cascade of events. It begins with an invitation to screening based upon risk—for average risk adults generally age and gender. It continues with the test itself. It is very important that that test be of high technical quality and that it be interpreted accurately. The results will be negative, positive, or indeterminate. Based upon those results, a timely follow-up in the near term or after the recommended screening interval will be required. Most women who undergo breast cancer screening will need repeat screening after a year or two years. In this series of events, there is tremendous potential for slippage, and failures at any one of the steps can nullify any gains or the high quality of any previous step and reduce the value of screening.

A successful screening program requires participation by a target population and health care providers and adherence to recommendations, especially the screening interval. Screening intervals are established based upon estimates of a detectable preclinical phase, the sojourn time. For all testing, we need to have adherence to quality assurance standards. For those women who have positive test results or even those women who have normal results, we need to have timely and thorough follow-up. Women who are diagnosed with cancer need to have state of the art treatment. Finally, we should have a comprehensive surveillance system to measure program performance, and we especially need feedback to participants. Participants include not only all the professionals involved in screening, diagnosis, and treatment, but it is also important that women hear how well we are doing in reducing deaths from breast cancer.

Let's begin with adherence to screening recommendations. In the United States screening is opportunistic, as contrasted with organized screening, which is more common in Europe. This means that it depends upon a coincidence of interest between providers and patients. The lack of population registries or reminder systems in the United States means that most American women do not get regular mammograms. Utilization of mammography is high, but utilization of regular mammography at appropriate intervals is not so high. For example, data from the New Mexico Mammography Project show that in that screening program 30 percent or fewer of women adhere to annual screening recommendations (Gilliland et al., 2000).

Failure to obtain regular mammograms is not harmful for the large majority of women who do not have breast cancer, but out-of-interval mammography does increase the risk of being diagnosed with advanced disease for those women who develop breast cancer during the missed interval.

Data over almost a ten year period from the Massachusetts General Hospital on a series of nearly 60,000 women and nearly 200,000 screening mammograms that resulted in the detection of 604 invasive cancers showed that tumor size was strongly associated with regular screening (Michaelson et al., 2002). Table 2.1 shows that at first screening, mean tumor size was 13.7 millimeters; on subse-

TABLE 2.1 Consequences of Screening Patterns for Breast Cancer Detection

Tumor Type	Number (Total = 810)	Mean Size (mm)
1st Screen	115	13.7
Subsequent screen	312	11.7
Never screened	206	15.0
Intervening (all)	179	16.8
Intervening (< 1 yr)	68	15.5
Intervening (> 1 yr)	111	17.6

quent screens in women getting regular screening, mean tumor size was smaller (11.7 mm). During the same time period, mean size of 206 invasive tumors in women who were never screened was a centimeter and a half. Intervening breast cancers, that is, cancers that were diagnosed at some point after a normal screening mammogram were 16.8 mm overall, but larger if the interval since last screening exceeded one year.

Fifty percent of women in this series did not have their first mammogram until after the age of 50, although 25 percent of the cancers were diagnosed in women under the age of 50. Among women screened for breast cancer, the majority did not return for repeat screening within a 12 to 14 month interval, and by a year and a half after their last mammogram, only 50 percent of women had returned. Twenty-five percent of breast cancers were found in women with no history of a prior mammogram. The median tumor size for these women was a centimeter and a half, compared to a centimeter in women attending regular screening (not shown in Table 2.1). Thirty percent of breast cancers were not found on mammography, and these were also larger than those found on women who followed screening recommendations.

As the data on these intervening cancers in Figure 2.8 show, only three percent were found in the first six months after the last normal mammogram, and only nine percent were found between 6 and 12 months. So just a little over 1 in 10 of these were found in the first 12 months after screening, what we would call interval cancers. Based upon the estimates of doubling time, the majority of these tumors emerged as larger palpable masses, not because they were missed, but because simply too much time had elapsed since the last normal mammogram. In fact, Michaelson and colleagues concluded that since so many of these women did not adhere to screening recommendations, almost 50 percent of the invasive tumors were larger and, thus, potentially more lethal.

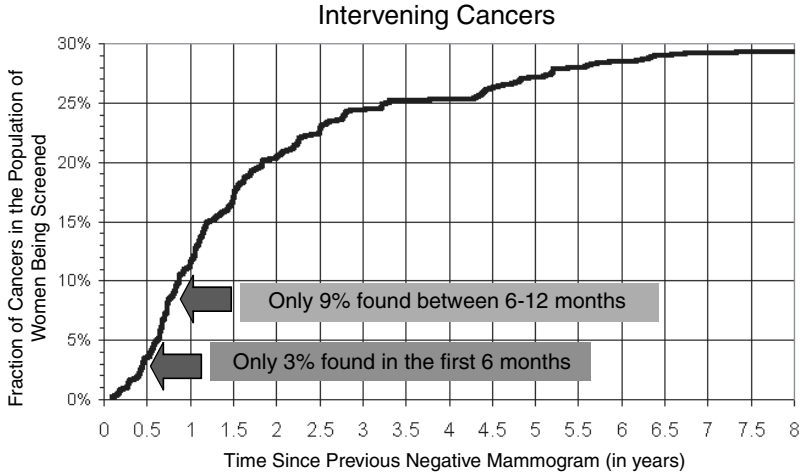

FIGURE 2.8 Breast cancer found after mammography.
SOURCE: Michaelson et al., 2002.

As noted earlier, opportunistic, that is, encounter—not population—based screening is one of the reasons for poor adherence to mammography screening recommendations. The situational context of encounters, then, is a limiting factor. Box 2.3 summarizes the problems with this approach to preventive health.

BOX 2.3
Deficiencies in Opportunistic Screening

Opportunistic (i.e., coincidental) preventive care is inherently limited:

- Encounter based, not population based
- Situational context of encounter is a limiting factor
- Depends on M.D. (preoccupation, forgetfulness, lack of familiarity with recommendations, or non-evidence based policy)
- Partial adherence is more likely than complete adherence
- More complex situations (follow-up, greater individual risk, etc.) are less likely to be properly addressed

In contrast, organized screening relies on a systematization of care, that is, institutional policies, population registries, computerization, reminders, chart tools, and audits to meet standards of care. Screening in this context, then, is less dependent on encounters. Outreach is easily modified for individuals and populations. There are fewer demands on providers. The systems for invitations, appointments and follow-up can be integrated, and there is the built-in potential for evaluation which is more efficient and less expensive than chart audits.

Compared to the systems in Europe, our non-system, or aggregate collection of unconnected small systems, is inherently complicating for developing organized screening. Could it be individually based? In other words, should individuals keep their own reminder systems? Some Internet systems provide a calendar that can notify a woman when she is due for screening or some other encounter, or she can use a small booklet, such as those provided by Putting Prevention into Practice (Reynolds, T., 1999). Should systems be office-based where both chart reminders or office computer systems are possible? Could we construct a central program based on population registries? We have population registries, but they are typically not used for health. Could the state health department take responsibility? Or could a health plan or a consortium of health plans unite together around some common goals of tracking and notification for preventive care?

Different outreach models have shown different degrees of success. Encounter based systems can improve cancer screening rates, but they are inherently limited because patients must initiate encounters. If a woman does not return to her doctor in a year or two or three, the physician will not be looking at the chart and seeing that that patient needs screening. Continuity with providers and practices is a major problem today as people cycle in and out of health plans. With increasing age, encounters are more likely to be chronic visits versus preventive visits. Only about one in four adults over the age of 40 gets a regular checkup.

Preventive service office systems depend on establishing practice routines, tools such as flow sheets, or defined responsibilities among clinicians. Paper based systems are effective, but computers measurably increase productivity. Unfortunately, just because you build it doesn't mean it gets used. The literature documents high failure rates among these systems; oftentimes they work very well when first initiated, and then their efficiency decreases.

A meta-analysis of 108 studies of strategies to increase rates of adult immunization and cancer screening through interventions including reminders, organizational change, feedback, education, financial incentives, legislative change, mass media, and even separate preventive care clinics found that organizational change was most effective in improving rates of preventive care. Among such effective changes were use of separate clinics devoted to prevention, use of a

planned care visit for prevention, and designation of non-physician staff to do specific prevention activities (Stone et al., 2002). Financial system incentives and reminder systems run second in effectiveness, and patient education, of course, was the least effective intervention. It should not be surprising that dedicating a time, place, and staff to preventive health is a more successful strategy than attempting to achieve some preventive health goals during encounters for acute and chronic conditions.

To summarize, both office systems and centralized systems have been shown to improve use of preventive care. When comparing an outreach system with an encounter system, patient-initiated encounters are the major limiting factor. Chart reminders and chart audits are less effective and less cost-effective than computerized outreach systems, and centralized systems provide for continuity and reduced stress on the practice and the individual provider. However, they do not eliminate the need for office routines and policy.

Given the way health care is not organized in the United States, single disease interventions have greater short- versus long-term potential; these programs, such as single disease computer software tools, generally may show some benefit, but inherently they will be less attractive to patients and providers than a comprehensive system. In addition, one reasonable strategy for breast cancer, since I think that a centralized system is relatively hopeless at the moment, might be simply to encourage radiology departments to manage the call-recall system.

Ultimately, a more organized approach to breast cancer screening would monitor population-based access. It would improve standards of screening based on evidence. It would monitor performance in terms of detection of small cancers, and it would implement technological improvements in early detection. However, there is no organization in the United States charged with ensuring that the availability of mammography is adequate to meet screening needs. There is no organization charged with ensuring that American women have access to mammography.

As shown in Figure 2.9, there are over a thousand fewer mammography facilities today than existed in 1994, yet we do not know whether this represents a consolidation of facilities and actually greater efficiency, or whether it represents markedly less access. We know that the decline has been greater in rural areas, so it is reasonable to suspect that rural women may have less access to mammography than they had previously. We also know that it is hard to reconcile a decline in facilities and an increase in units. Capacity at some of these facilities is clearly increasing, but we have no idea whether it is increasing enough.

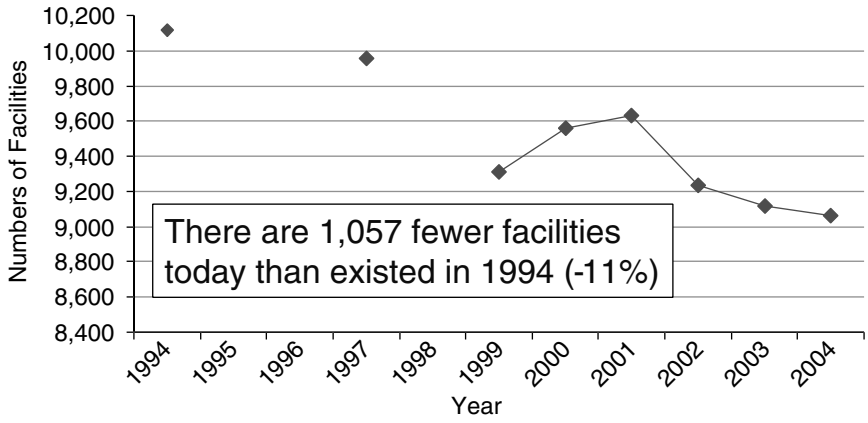

FIGURE 2.9 Declining mammography facilities.
SOURCE: FDA data.

The GAO report on utilization (GAO, 2002) estimated that, assuming 4,500 exams per 10,000 women per year, there should be approximately 2.2 mammography machines for every 10,000 women in the population if around 90 percent compliance with screening recommendations is the objective. Only five states have a ratio of 2.2 machines per 10,000 women, and 11 states have a ratio of two per 10,000 women, so it seems that the majority of states do not have the capacity to deliver recommended services at the 90 percent adherence rate.

According to an American Hospital Association survey, in 2001 the job vacancy rate for technologists was 18 percent, and almost two-thirds of hospitals reported difficulty recruiting them. Also, fewer technologists were seeking mammography certification. The attractive jobs for radiologists are also the attractive jobs for technologists, and the unattractive jobs for radiologists are also quite unattractive to technologists. So, as mammography has become less appealing to radiologists, it has also become less appealing to technologists for many of the same reasons.

In a telephone survey that examined radiology residents' attitudes towards going into mammography, 63 percent said they would not like to spend 25 or more percent of their time in practice interpreting mammograms. The most common reasons for avoiding mammography were that it was not an interesting field, the risk of litigation was too high, and it was too stressful. Most residents (64 percent) reported that they would not consider a fellowship in breast imaging if offered (Bassett et al., 2003). As noted earlier, mammography is also unappealing to radiologists because earnings are comparatively lower, the financial

contribution to the practice is small, and mammographers, therefore, are not as appreciated as some other members of a radiology practice. Work load, stress, and malpractice exposure are high. The procedure is repetitious and low tech, and it is not highly respected by colleagues.

I think each of these complaints can be addressed. We can certainly do something about earnings and the financial contribution to a practice. If we perform mammography more efficiently, and we raise the reimbursement rate, then the contribution is going to be higher. There are a number of ways that we can build solutions for the work load, both for technologists and for radiologists. There are ways that we can reduce stress. Malpractice exposure is high, but, personally, I do not think tort reform is going to solve this problem. I think we need more creative solutions, including thinking very seriously about a no-fault system, much like we have for vaccines. The procedure is repetitious and low tech, but higher tech imaging is coming along. The consideration of physician extenders is also quite reasonable, but here again the threat of increasing malpractice exposure for a supervising physician is a real obstacle to expanding the workforce to include non-physician interpreters. Again, I think the problems can be addressed. Of course, we should not expect all radiologists to want to do mammography. We just have to be sure that there are enough.

In thinking about what we might achieve through high quality, the differences between an unorganized and an organized health care delivery system provide some useful examples. Overall, however, as shown by some of the receiver operating characteristic data recently reported (Beam et al., 2003a), the quality of mammography in the U.S. is good, but it is quite variable. I think that is where the greatest concern lies.

In British Columbia, radiologists who read mammograms must pass an exam to show that they are proficient at finding small cancers. There is a strong interest in reading mammograms in British Columbia, because it represents extra income, and it is an environment where there is not a lot of competition from high tech procedures, certainly not compared with the U.S. The statistics for their program are really quite good as shown in Table 2.2 (British Columbia Cancer Agency, 2003).

I also want to point out the effect of age. The positive predictive value, that is, the proportion of women with an abnormal mammogram that actually have breast cancer, is lower in younger women, but these numbers improve as women get older. Across the board, the median tumor size is quite small, and the percent of node-negative tumors approaches 3 out of 4. This is an example of what ahigh performance program can deliver, and so I would encourage you to think not so much in terms of what our current data about mammography tell us is achievable, but more in terms of what the data tell us we might achieve through organized high performance systems. This should be our objective.

TABLE 2.2 A High Performance Mammography Program in British Columbia

	Age Group		
	40-49	50-59	60-69
No. exams	81,515	68,746	44,457
Abnormal	7.6%	7.0%	6.1%
Overall Cancer Detection Rate	1.7×10^3	3.9×10^3	5.6×10^3
PPV	2.4%	5.8%	9.7%
Median Tumor Size	13mm	12mm	13mm
%Node-negative	.70	.73	.73

SOURCE: British Columbia Cancer Agency.

New York State has taken an oversight responsibility for the 292 facilities that participate in their state's Breast and Cervical Cancer Screening Program funded through the CDC National Breast and Cervical Cancer Early Detection Program (Hutton et al., 2004). The State Health Department regularly scans the data looking for outliers that exceed upper or lower bounds established for the program. They monitor the abnormal clinical breast exam rate, the abnormal mammography rate, the positive predictive values for biopsy, the cancer detection rate, and age, race, ethnicity and previous screening history among the participants. They visit facilities that have outliers, evaluate them, review medical records and mammograms for quality and interpretation, and they review their breast exam techniques.

Since the average radiologist, and even the average facility, detect few cancers in any given year, some surveillance measures may be strongly influenced by chance or other factors that could mistakenly portray a facility as performing very well or very poorly. Thus, this kind of program should not be, and is not punitive. Rather it is a responsible surveillance program designed to monitor performance and provide feedback, which, as I mentioned at the beginning of my presentation, is critically important to maintaining a high degree of quality assurance. That is why you have a visit. That is why you do surveillance and reviews. Following the data review, if corrective action is necessary, the state collaborates with the facility and the local screening project to implement corrective actions.

TABLE 2.3 An Outlier Mammography Facility Before and After a Collaborative Corrective Program with New York State

Indicator	Facility A	After Corrective Action
No. Mammograms	544	-
Abnormal	166(31%)	-
No. BI-RADS 4	73(13.4%)	4.3%
Additional Imaging	5/166(3%)	7%
Biopsy Rate in Abnormals	52%	28%
PPV_b	6.9%	19%

Table 2.3 provides an example that compares statistics from a facility that was identified as an outlier to the same measures from that facility after a visit and a plan of correction. The facility had relatively low volume, but it had a high abnormal mammogram rate, a relatively high rate of BI-RADS® 4 interpretations, a very low rate of additional investigational imaging, a high biopsy rate of women with abnormal mammograms, and a relatively low positive predictive value on biopsy.

During the state's visit, it was discovered that two staff radiologists had higher biopsy rates than the other three. In-service training for all five radiologists on the use of the BI-RADS® system and double readings for six months were instituted, and the facility was encouraged to obtain accreditation for its stereotactic biopsy program. After corrective action, the number of BI-RADS® 4 readings dropped to 4.3 percent; additional imaging to reconcile abnormalities more than doubled; biopsy rates in women after abnormal mammograms declined substantially, and the positive predictive value for biopsies increased.

This sensible surveillance strategy can improve the quality of breast imaging in ways that are difficult for voluntary programs or the Mammography Quality Standards Act to achieve, and it also can detect fraud and dangerously low quality. For example, the state discovered a facility that was reporting clinical breast examinations that were actually not being done and determined that the radiologists interpreting films were doing such a poor job that they shut the facility down, and the American College of Radiology withdrew its accreditation.

I turn now to a more detailed look at the advantages and disadvantages of organized screening. Organization can lead to more formal and uniform decisions about whether, whom, and at what intervals to screen. It also can install or improve uniform call-recall systems and triage, improve the timeliness of follow-up, minimize loss to follow-up, and improve quality assurance, monitoring, and evaluation.

It is important to understand also what organization may not accomplish. First of all, resource issues may take precedence over evidence, both nationally and locally. For example, screening programs in Europe may have elements, such as the recommended 3-year screening interval in the United Kingdom, that are due to resource limitations and are not in keeping with available evidence. Poor or incomplete population registries may limit the effectiveness of any call-recall system.

The screening program may be under funded; I cannot think of a single program in Europe that isn't stressed by lack of funds. Resource limitations may require various compromises that influence program goals, such as balancing sensitivity, specificity, and positive predictive value, or unequal access, or less attention to overcoming barriers. There simply may not be resources for outreach to hard to reach groups, even though the commitment is there in principle. And there may be delays in acquiring new technologies. For example, I know that there is quite a bit more digital mammography in the United States than in Europe at present.

Opportunistic screening may be more appealing to some groups in the target population. We have observed that it is sometimes very difficult to impose organized screening on a population that has always had a choice of where to go and whom to see for screening. We observed in some European countries that some providers were not enthusiastic about the organized program and participation rates by women in those programs were not as high as elsewhere. At times, participation in organized screening may not exceed that in opportunistic screening. Participation varies in both models by gender, socioeconomic status, perception of risk, rural-urban status, and various attitudes, and is quite a bit higher in the United States than in some organized settings in Europe. Organized screening does not eliminate the need for behavioral interventions over time focused on key target groups, such as groups with low income, groups that are institutionally adverse, and individuals who tend to refuse screening.

Nevertheless, based upon what we observe today and upon how much more we might achieve, I believe we must think about delivering organized breast cancer screening in this country. Organized breast cancer screening would increase adherence to regular screening. It would improve accuracy, and we would see an increase in the rate of cancers diagnosed before they become advanced. Data from Dalarna County in Sweden over a period of 40 years, shown in

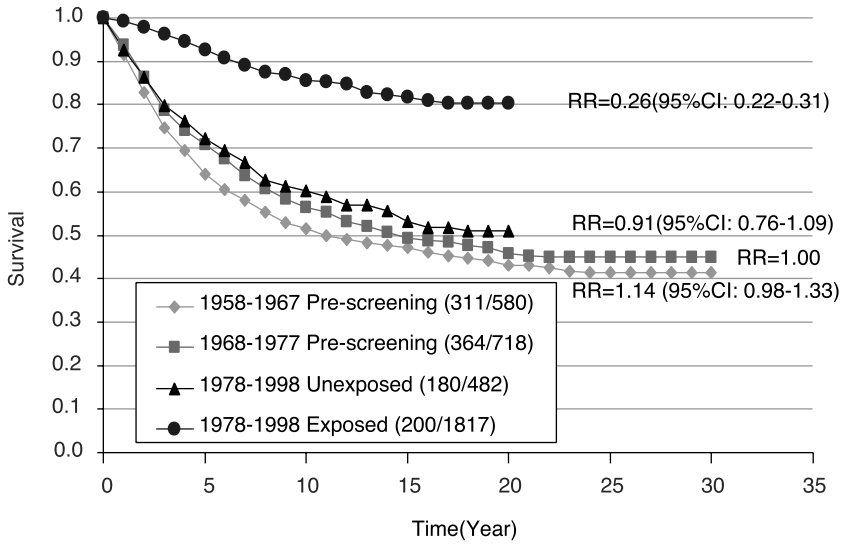

FIGURE 2.10 Dramatic improvements in survival in breast cancer in populations with high-quality organized screening: cumulative survival of breast cancer patients age 40-69 in Dalarna, Sweden. Diagnosis and death 1958-1998. SOURCE: Tabar et al., 2003

Figure 2.10, document survivals over 20 or more years in groups of women from two prescreening periods, 1958-1967 and 1968-1977, and compare those to survivals of two populations from the screening period (1978-1998), those who were exposed to screening and those who were not exposed to, or refused, screening. In the modern, screening period, survival is strikingly better for women participating in a program that has very high rates of adherence and very high quality (Tabar et al., 2003).

We really should be striving to do the very best we can in mammography, and doing our best means devoting attention to increasing benefits and reducing harms associated with screening.Ultimately our goal for mammography must be saving women's lives by detecting breast cancers before they become advanced, when the patient has the best chance for successful treatment.

DR. PENHOET: We have time now for a question and answer period for the three speakers of the morning.

DR. SMITH-BINDMAN, University of California, San Francisco: I am inspired by the possibility of having a single organized screening program. I ask both Dr. Dunnick and Dr. Esserman about the feasibility of an organized

system from their perspectives. Dr. Dunnick, you presented a very bleak picture of the way mammography is organized now. Financially it is not working at all; I think reimbursements in the academic environment are actually lower than in private practice, and there are a lot of impediments to building a better system. One solution is to throw more money at the system. I am not optimistic about that approach, because the current system is inefficient. Do you see as a possibility having freestanding organized screening programs as a way to give some status to radiologists who do mammography, to increase the finances, to increase the efficiency? Would that be a solution to the current broken system?

DR. DUNNICK: It certainly might help. Of course, as always, the devil is in the details. We need to know exactly what is proposed rather than just a general suggestion, then we would be able to assess that more easily. Mammography, of course, is not the only area that is relatively under-reimbursed. You really could look at the whole spectrum and try to align the reimbursement more closely with the work effort. We also have to support areas in nuclear medicine, pediatrics, as well as mammography.

DR. SMITH-BINDMAN: But sticking to mammography, I have trouble with the concept of mammography being under reimbursed. Regarding diagnostic mammography, clearly there is a lot more work that goes into it, probably tenfold more work than screening. But screening mammography could be very efficient in the current reimbursement scheme if radiologists didn't hang films, call reports, that is, basically do clerical jobs. In the study that is widely cited, the radiologists only spent a third of their time doing radiology (Enzmann et al., 2001). So the current $40 reimbursement for the professional component of mammography could be an effective reimbursement to a good mammographer who could interpret a mammogram in about a minute if all the clerical diversions were eliminated.

DR. DUNNICK: We might not have any good mammographers at Michigan then because a minute seems rather quick to me.[2]

DR. ESSERMAN: I think that is a very good question. There are opportunities to learn from how breast cancer screening is organized in Sweden and the U.K. Prior to about 1989 or 1990, there was no systematic screening program in the U.K. Obviously it is easier for them to implement changes because they have a centralized organization, but their screening program was not integrated into the National Health Service; it was set up separately. I am encouraged by Dr. Smith's discussion of what is happening in New York showing that there are other ways to impose organization and to set up a program with benchmarks.

---

[2] Mean time for screen film mammography reading was 79.5 seconds (range, 15.8-444.8 seconds) and for digital mammography reading was 159.16 seconds. (range, 23.0-587.1 seconds) in a study at the Department of Radiology, Michigan State University, by Aben GR, Bryson, HA, Bryson TC, presented at the International Workshop in Digital Mammography, Chapel Hill, NC (personal communication by Etta Pisano, 2004).

I like the idea of thinking about opportunistic versus systematic. A reduction in breast cancer mortality rates is one of the benefits of the screening program in the United Kingdom, but it was not achieved by screening alone. Prior to the program, 90 percent of all breast cancer patients were cared for by general practitioners. Within 7 years of instituting the screening program, 95 percent of all women were treated in organized breast centers. That is not the case today in the U.S. So by taking an organized approach and looking at things systematically, you can make change. Yes, it is harder in our current climate, but I don't think it is impossible.

I think there are organizational examples of ways to leverage the time of the radiologists considerably. There are important ways to leverage biology as well. Some of the work towards risk stratification schemes may ultimately allow us to focus more on diagnostic mammograms and screen in a smaller group of patients, which could improve efficiency and solve some of the manpower issues. What is helpful about this report is its challenges to think about population registries and state systems to start some systematic programs.

DR. SMITH-BINDMAN: I had a second question directed to you regarding selective screening, which I think is a fantastic idea. We obviously don't have the tools yet to assess risk. We have age and breast density as a risk, but proteomics is in the future. I am concerned that if we have no system new tools will be embraced rapidly, and, rather than leading to selective screening, we will subject women to all of the tests, which is what we do in medicine, as opposed to thinking about how to use things efficiently.

DR. ESSERMAN: The committee members agree with you. Recommendation C proposes integrated technology centers to help design and evaluate systems. As Dr. Smith said, screening is a cascade, integrating the testing with the treatment approach. Understanding the biology allows us to avoid over-treating as well as under-treating. Programs like the NCI's Early Detection Research Network could be leveraged to help build centers around testing, developing, and deploying. People like Dr. Bohmer of the committee have noted that we are going to have to get to work on both implementing systems for population-based screening and for developing and deploying new technology properly.

DR. WARRICK, University of Texas School of Public Health: Dr. Esserman, what do you mean by regionalized programs and, in terms of understanding risk, are you suggesting that we abandon the Gail model (Gail et al., 1989)?

DR. ESSERMAN: By regionalized programs, I mean what Dr. Smith was describing, not having an opportunistic, but rather a systematic approach. For example, the U.K. program takes the population-based registry for a region, sends letters to all women in the screening age range who are registered, and invites them to screen. There is an organized system that tracks outcomes and monitors benchmarks. Instead of requiring extra work as in New York State,

these things are part of the routine. That is an organized approach, and it is why systems are critical to realizing the full benefits of screening mammography.

In terms of the Gail model, no, the Gail risk model is not out. At a risk models meeting here a few weeks ago, integration of the approach to prevention and screening with understanding underlying risk was discussed as a way to tailor screening strategies. I think all of the risk models need to be integrated and new biologic tools added to better tailor screening. The Gail model has value in understanding that a five-year risk and a lifetime risk can actually be very important, and might be used as one of the screening strategies.

DR. SMITH: The value of some of these risk models lies in identifying women that are at measurably higher risk. Identifying a woman at low risk still obliges you to ensure that a breast cancer is detected this year or next year. Ultimately, even a woman who is at the lowest risk in the Gail model is still at appreciably enough risk to require screening. So, these models may be of value in identifying women who need a different screening interval, to begin at an earlier age, to undergo more frequent screening, something that is tailored to her uniquely. Even in some countries with a third the risk of the United States, there is interest in exploring how to prevent a late-stage diagnosis.

DR. PENHOET: Perhaps new technology will expand the definition of each of the four components in the Gail model. One of them is family history. Modern genetics should allow you to explore family history in much more precise genetic terms than simply as a broad category. With respect to the other factors as well, a deeper understanding could expand each one. So I don't think anybody would argue that the Gail model is not relevant, but there will be additional subcomponents under each of those four categories that are listed there today as the likely outcome, I think.

DR. TAPLIN, National Cancer Institute: Later in this symposium, I will be talking about an organized screening program here in the U.S. beginning back in the 1980s. So there are examples of organized screening from the past in the United States. It is an entirely possible model, but it requires leadership, oversight, and some commitment to an explicit set of guidelines.

Looking at the organized European programs with leadership and explicit guidelines, most of these have a two year screening interval. One of the underlying issues that you have raised today is capacity. The math is easy. If we went to a 2-year interval, especially for women 50 and above, you would automatically create more capacity. I ask how the panel feels about that, and why that isn't considered a possibility for addressing capacity and access problems in the U.S.

DR. ESSERMAN: I think that is a great comment. It may be that you can tailor that approach by age. The choice that was made in the U.K. was to screen every 3 years, not because they thought that was the optimal interval—they actually thought 2 years was the best interval for screening at ages 50 to 65. But they only had 128,000 pounds sterling annually. They made trade-offs because they decided that putting in a systematic and organized program where they

could track outcomes would be their best first investment. In Sweden, they also have done a lot of work in looking at different intervals.

I think your point about capacity is very important. The opportunity to integrate risk models, to understand what kind of cancer you are at risk for, what the right interval should be, to understand the biology of disease, might allow us to tailor the interval much better in the future and to fit capacity to our need.

DR. SMITH: The screening interval should be driven by the estimated sojourn time. Therefore, screening every 2 years for women between ages 40 and 54 would not be advisable. Although we previously recommended to women in this age group that they be screened every 1 to 2 years, it seems the wrong message. Epidemiologists looked at the trials; some screened every year, some every 2 years; there was a benefit at each interval, but not the same benefit.

The trials and the Swedish experience have shown the need to tailor the screening interval to the detectable preclinical phase. We recommend annual screening for women over the age of 50 because that is what the HIP trial did. As we have learned more about sojourn time, we understand that nothing happens abruptly at the age of 50. Probably as women approach the age of 60—and this is what the Swedes did—one might extend the screening interval. This was a function of resources in Sweden. The data from San Francisco show better performance with annual screening compared to biannual screening in postmenopausal women although the improvement is less than it is in premenopausal women (Hunt et al., 1999). So if you can learn more about risk, for example, if a woman is not on hormone replacement therapy, has large breasts, and her mammograms are easy to read, then after the age of 60 it might be reasonable to screen her every 2 years.

DR. TAPLIN: I think this is a question for the panel to consider. We know that the sojourn time for women ages 50 and above is quite long; 3 years is a common estimate. So a 2-year screening interval is something to consider as a way of increasing capacity and access.

DR. ESSERMAN: It also seems that discontinuing hormone replacement therapy in women over 70 makes more of an impact than screening. So again, it is changing your underlying risk that can help you determine what the right intervals are.

DR. SMITH: But post-menopausal women with a family history do not eventually graduate to the longer sojourn time. These are women that really do need to continue to be screened annually.

DR. ESSERMAN: And if you have a systematic program, you will be tracking outcomes and looking at the data; you can look at variation in care regionally and at interval cancer rates. Because we don't have an organized approach, we are not collecting information routinely and systematically, and we can't answer those questions.

MRS. LAUDER: Given that there are fewer radiologists and breast imagers because of insurance reimbursement, liability, and lawsuits, are airline pilot restrictions on work loads an example of a strategy to improve efficiency and safety that might be applicable? Are there standards that inform radiologists about a limitation on the number of mammograms that they should interpret? Have any eye doctors come up with any recommendations as to what is physically possible? The idea would be to help in reducing the risk of human error, increasing efficiency, protecting the physician from lawsuits, and getting more accurate readings for the patients. I'm looking for ways to make mammography more attractive to radiologists.

DR. DUNNICK: That is an interesting question, and I don't have the answer. To put it another way, at the end of the day does the radiologist miss more cases than at the beginning of the day? Is there an increase in the error rate as one tries to become faster and read more cases? I don't actually know the numbers.

DR. SMITH: I think some of our colleagues in the audience might be able to address this. There is a literature on this that I have only begun to see, some of which evaluated how people tried to interpret other radiographs not just mammograms, the patterns of eye movements and the like. It would be very interesting to look at fatigue in the context of training and competence. Clearly a radiologist who was not comfortable reading mammograms will be reading under duress and may spend more time and experience fatigue due to uncertainty and anxiety. In addition, there will be the stress caused by the importance to women of detecting breast cancer and the threat of litigation for missed diagnoses. Of course, no matter what the level of competence, there will be fatigue at some point, but I think that radiologists who specialize in this field know when they are tired. They know when to stop reading and take a break, or how many films they can interpret during the day. We have done studies with cytotechnologists, to evaluate their patterns of reading pap smears. That led to limits under regulations of the Clinical Laboratory Improvement Act.

DR. BORGSTEDE, American College of Radiology: The question is an excellent one. A recent study analyzed the true to false positive ratios for numbers of mammograms read. The optimal was between 2,000 and 4,000. Radiologists reading fewer than 2,000 or more than 4,000 had poorer ratios (Kan et al., 2000).

DR. D'ORSI, Emory University: Numbers are not the only factor in adequate interpretation. You can read any number of mammograms and read them wrong constantly. So, along with numbers you have to review your results. Screening is a test that has built-in high accuracy because 99.5 percent of these exams are negative. If you simply read them all as negative, your accuracy will be 99.5 percent. The task is to pick up malignancy. False positives are the currency that you pay to detect subtle malignancy; as you increase false positives, your false negatives decrease. Obviously, there is a point where this is no longer

productive. But before we dwell too much on the undesirability of false positives, we have to think what those false positives are getting for us. With good intelligent readers, they are getting us subtle malignancies.

DR. ESSERMAN: Alistair Gail, a Ph.D in visual psychology, has developed the U.K.'s performance test. This is a series of mammograms required as a training set for every mammographer. The program has organized special training sessions and modules based on the kinds of things that they miss. We should keep in mind that providing training and feedback are among the important factors in improving mammography.

Risk Stratification for Breast Cancer Detection: Better Quality Mammography for Women Through Better Focusing of Services

Suzanne Fletcher, M.D., Professor of Ambulatory Care and Prevention, Harvard Medical School

I first want to congratulate the committee that produced this report. As a member of the committee that produced the previous report, *Mammography and Beyond* (Institute of Medicine, 2001), I have some idea of how hard you worked.

In my presentation today, I will discuss the science of risk stratification, breast cancer risk in context, and breast cancer risk perception. I will spend most of my time on the first topic.

The report lays out a hope for risk stratification. The idea is to take all women and assign them to groups that have ultra-low risk, medium risk, and high risk for developing breast cancer. There are many advantages to such a plan. From the individual woman's perspective, those that are low risk may be reassured, and they may also require less screening to protect themselves from breast cancer. For those at high risk, although it is disturbing news, they at least are informed and can take steps to prevent harm. From society's perspective, concentrating efforts on women in whom we can prevent most of the adverse effects of breast cancer would be wonderful.

As I delved more deeply into risk stratification, however, I began to think it may be more difficult to achieve than I, at least, previously thought.

Let's start with some risk factors which the committee summarized in tables in the report. I have listed in Box 2.4 six major risk factors. By major, I mean the relative risk is at least three, that is, the risk of breast cancer in women with these factors is at least 300 percent that of the risk of women without these characteristics.

BOX 2.4
Major Risk Factors for Breast Cancer with Relative Risk at Least 3 or Greater

Major (RR.3.0)	
• Increasing age	~80(70 vs 30)
• Genetic Mutation	~200(<40)
	~15(60s)
• Atypical hyperplasia	~5
• Radiation therapy	~5
• Omcreased breast density	~4
• Strong family history	~3-4

Increasing age is a well-known risk factor, one which we already use for breast cancer screening. The size of this risk factor depends on the cut points; I have compared women in their early seventies versus women in their early thirties; breast cancer is about 18 times more likely in the older women.

I had not considered genetic mutation as big a risk as the committee reported. However, women under 40 with a deleterious BRCA-1 mutation are 200 times more likely to develop breast cancer than women of similar age without deleterious mutations. The relative risk of the mutation decreases as women age, but is still high, 15, in women from ages 60 to 69 (Singletary, 2003).

The other risk factors, although called major, actually confer far less risk, including: atypical hyperplasia; radiation therapy (for things like Hodgkin's disease, so that is not going to be relevant to most women); increased breast density; and strong family history. Box 2.5 lists many of the risk factors that

BOX 2.5
Moderate Risk Factors for Breast Cancer Moderate Risk Factors for Breast Cancer with Relative Risks 1.0-3.0

Mother or sister with breast cancer
Increased bone density
Older age at first birth
Older age at menopause
Younger age at mearche
Benign breast biopsy
Alcohol consumption
HRT/Contraceptive pills

women often think about and that are frequently discussed in magazine articles; they are actually rather modest, with relative risks under three, and except for family history, they all may be related to exposure to estrogen over time.

For risk stratification, the idea is to use these characteristics to stratify women into low- and high-risk groups. The best-known example we have of a risk stratification tool for breast cancer is that developed by Mitchell Gail and colleagues (Gail et al., 1989, and see NCI website http://bcra.nci.nih.gov/brc/). This tool was developed in 1989 from the Breast Cancer Detection and Demonstration Project based on risk factor information from about 200,000 women and was used to estimate expected breast cancer incidence in the Breast Cancer Prevention Trial. The risk factors in the model, which applies only to women over age 35 and assumes regular screening, include age, age at menarche, age at first birth or nulliparity, number of affected female first-degree relatives up to two, and history of benign breast biopsy or hyperplasia. Missing from this list, however, are major risk factors like breast density or mutations.

How well does this tool work in stratifying women according to their breast cancer risk? A recent report examines this question (Rockhill et al., 2003). The investigators studied a cohort of 82,109 women ages 45 to 71 who were part of the Nurses' Health Study. They followed the women for the period 1992 to 1997, having collected information on their risk factors for breast cancer. Of the 82,109 women, 1,354 developed breast cancer during the study period (1.65 percent). Using the Gail model, risk was estimated for each woman. All these risks were summed, and for the group the model was found to work remarkably well. The ratio of expected to observed cancers was really excellent (0.94) and even better in a high risk subsample (1.03). However, the Gail model did not work as well for individual women, and that is what is needed for risk stratification.

An individual woman either does or does not develop breast cancer, so her risk is either zero or 100 percent. It is only in groups of women that you get a range from zero to 100. Because it was not clear to me how the investigators assessed risks for individual women, I consulted Dr. Rockhill. I learned that she and her colleagues evaluated the accuracy of the Gail model in two different ways. One way was calibration of the model as I have just described, that is, determining the degree to which the percentage of a population actually developing disease is similar to the probability estimate of the model for that population. The Gail model estimated 1.55 percent of women would get breast cancer in this population of women from the Nurses' Health Study, and 1.65 percent actually developed breast cancer. So the expected to observed ratio was 0.94; that is the calibration of the model for the group.

Stratification requires a model to go one step further—to discriminate. Discrimination is the degree to which the estimated probabilities from the model are consistently higher for persons who develop disease compared to those who do

not. Do the women who get breast cancer have a higher risk according to the model than the women who do not? This is calculated according to a concordance statistic, the values of which run between 0.5 (a coin-flip) and 1 (perfect discrimination). One woman who got breast cancer and one woman who did not get breast cancer are randomly selected, and it is determined whether the woman who got breast cancer had a higher risk score in the model than the woman who did not. All these determinations are summed, and the result is a percentage which represents the probability that for any randomly selected diseased/non-diseased pair of women the diseased woman has a higher estimated risk.

In the Nurses' Health Study, the resulting percentage was 58 percent. Now, 50 percent is flipping a coin; 58 percent is better, but let's face it, it is not much better.

Another way to demonstrate the concordance of 58 percent is with a graph. Figure 2.11 shows the percentage of women in the two groups with different estimated five-year risks according to the Gail model. Effective stratification should separate these groups. The Gail model did not separate the two groups. There is no place along the horizontal axis to draw a line, above which we would offer screening and below which we could reassure individual women. Also it is important to remember that the figure shows proportions of women, not absolute numbers. Almost 80,000 women did not develop breast cancer in this 5-year period, whereas 1,354 did. If the figure displayed absolute numbers instead of percentages, the group that did not develop breast cancer would swallow up the group that did.

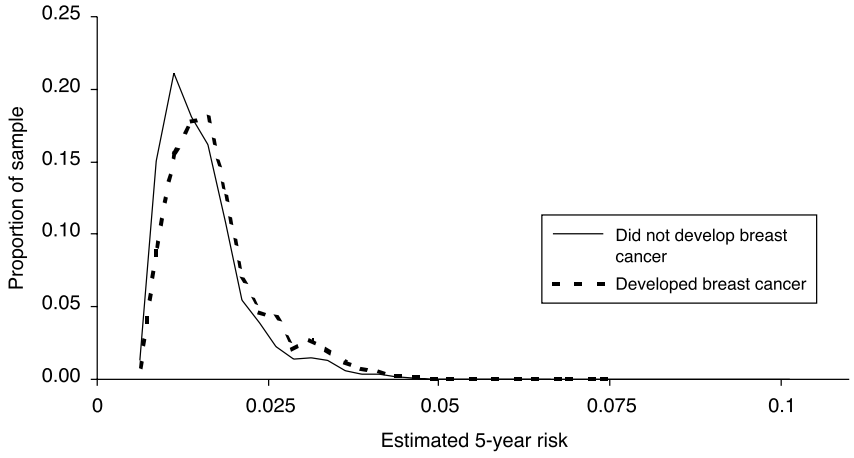

FIGURE 2.11 Discrimination of the Gail model.
SOURCE: Rockhill, B, personal communication, reproduced with permission.

For several other medical conditions, we know that a large number of people at low risk may give rise to more cases of disease than a small number who are at high risk. This common situation seems to be true for breast cancer, and it limits the utility of stratifying women according to risk when considering breast cancer prevention strategies. This is a reality that we must understand as we think about risk stratification.

How big does a risk have to be to be useful for stratification? In the Gail model, the relative risks were small. Except for age, no factors with very large risks were included. So how large a risk will be needed to discriminate between women who will get breast cancer and those who will not? A report in the *British Medical Journal* (Wald et al., 1999) suggested that a relative risk of about 200 is necessary to discriminate well between groups, and risks of this magnitude are rare; there are a few examples like alpha-fetoprotein and spina bifida (at 242) and perhaps hepatitis B and hepatocellular cancer (at around 200).

In breast cancer only BRCA-1 or 2 in young women are in this range. To test BRCA as a discriminator, I made a calculation using the relative risks I mentioned earlier of 200 for women ages 20 to 49 and 15 for women between ages 60 and 69. Although we are not really sure about these risks, they will serve for the purposes of this example. Also, although not certain, I used the estimate that about 0.25 percent of women carry a deleterious genetic mutation for BRCA-1.

Table 2.4 displays the figures from my "back of the envelope" calculation using the SEER data for the risk (and numbers of cases) in the younger and older general populations. In the young age group with fewer cancers, the high-risk BRCA positive women account for half the total (5,000 of 10,000) which is consistent with the estimate of Wald and colleagues that about half of cases in a

TABLE 2.4 Discriminating Potential of a High Relative Risk for Breast Cancer, BRCA 1, in Young and Old Populations: Back of Envelope Calculation

	Number of Women	Breast Cancer Risk/Year	Number of Breast Cancer/Year
Age <40			
General Population	40 million	1/4,000	10,000
BRCA1 Mutation (RR 200)	100,000	1/20	5,000
Age 60-69			
General Population	10 million	1/270	37,000
BRCA1 Mutation (RR 15)	25,000	15/270	1,4000

population can be accounted for by a group with a relative risk of 200. I emphasize that this is a very rough calculation. There are all sorts of subtleties that are not included, but it is sufficient to make the point for breast cancer.

Turning to older women, we have many more breast cancer cases in the general population because the underlying risk in the population is higher. The subset of the population with BRCA mutations accounts for less than four percent (1,400 of 37,000) of all breast cancers in this age group. It appears the risks that work for breast cancer risk stratification are ones with very large relative risks, or perhaps a combination of factors that add up to a very large relative risk.

In summary, risk stratification may be difficult because most risks for breast cancer are small and because many of these risk factors are spread out over the entire population. As Wald and colleagues point out, some of the risk factors are calculated by using the extremes of the population, but we have to apply them to the entire population, watering down the discriminatory effect. BRCA-1 may be an exception. Exciting developments in breast cancer research may lead to other examples where we can identify other small groups of women with large risks. Regardless of what happens, the take-home message is that risk models should be evaluated not only for their calibration, that is, how well they work for the whole group, but for their discrimination: how well they can separate individual women who are and are not going to develop diseases like breast cancer.

I now want to shift the discussion and talk about breast cancer risk in context. I was delighted that this was covered in the report. American women think that if they know anything at all about breast cancer risk, they know about one in eight women are going to get this disease. I was glad to hear from Dr. Smith that the American Cancer Society has decided that it was time to downplay that lifelong risk, mainly because too many women translate it into a short-term risk. It is, perhaps, one reason why so many women think that there is an epidemic of breast cancer.

In my own view, at least health providers should know the absolute risks displayed in Table 2.5. One of our problems has been that so much of the

TABLE 2.5 Absolute Risks Among Women of Developing and Dying of Breast Cancer in 10 Years

Age	Develop Breast Cancer	Die of Breast Cancer	Die of Any Cause
40	15	2	21
50	28	5	55
60	37	7	126
70	43	9	309
80	35	11	670

dicussion about breast cancer is in terms of relative risks, which sound so much more threatening. While it is true that breast cancer is the biggest cause of death in women in their forties, it only accounts for 10 percent of the very few deaths that occur at that age (2 breast cancer deaths per 1,000 women). But women also need to understand the increasing incidence of breast cancer with age, and the resulting need for screening as they grow older. They must also begin to understand that a breast cancer diagnosis is not a death sentence.

Finally, they have to have some sense of putting this information into context with the rest of their lives, and indeed, so do we. We are spending a whole day on breast cancer quite appropriately at this symposium, but I know as a general internist that women have many other complaints and concerns.

Finding appropriate methods to communicate these facts to women is not easy. When I was at the University of North Carolina, and it got to be known that I was involved in breast cancer screening and prevention, several college students in their early twenties would come to my practice worried about breast cancer. For the vast majority, I could not find anything that would suggest they were at increased risk. I would tell them that although it is true as a woman gets older breast cancer risk increases, the risk at your age is somewhere around one in 100,000. This did not seem to be reassuring; they had no context for the numbers. Finally, I began to say, "Look, a 70-year-old man has five times your chance of getting breast cancer in the next year." You could just see the light bulb go on. So we have to figure out a way effectively to communicate the risks of breast cancer to our patients.

That is the reason I was also delighted with the report's recognition that breast cancer risk perception is important. There is a great deal of fear out in the community about breast cancer. Years ago, colleagues and I conducted telephone surveys in two communities in North Carolina, and found that about a fifth to a quarter worried about breast cancer, about half feared finding it, almost three-quarters thought looking for it made women worry (Fletcher et al., 1993).

The report cites a survey (Black et al., 1995) of young women in their forties over-estimating their risk of dying of breast cancer twenty-fold, and their risk of getting it about six-fold. We clinicians have tended to think that this problem is not medical, and so we do not discuss it or address it very often with our patients. Thank goodness, the committee thought risk perception is important and made recommendations to address it.

What is the cultural context of all of this? The ancients thought of the breast as nurturing, as being very important for the survival of the species. They made deities out of women who were able to nurture twins. They got so excited about the breast that in 2000 BC, they gave one of their goddesses about 20 of them on her chest. Then over the centuries, society's connotation of the breast took on more of a sexual importance. Only in the last part of the twentieth century, have we begun to think of the breast as also signifying death and mutilation. I think

there is too much of that sense in the perception of breast cancer risk in the United States today. Science and medicine obviously are not totally responsible for this perception, nor can we change cultural attitudes completely. But we should accept some responsibility for the modern fear of death and mutilation from breast cancer. It is time to work hard to give women a more realistic understanding of their risk of breast cancer.

In conclusion, the committee made two major recommendations about breast cancer risk. The first was to develop individually tailored risk prediction tools to identify women who would benefit from individualized approaches to breast cancer detection. I agree with this important goal, but I have suggested it may be more difficult than we thought. Problems worthy of attack prove their worth by fighting back. Maybe this is such a case. The other major recommendation was to develop tools that facilitate communication regarding breast cancer risk. This too may be difficult, but is very worthwhile.

The Promise of Biomarkers in Early Detection of Breast Cancer: Better Quality Mammography Through Better Focusing of Services

Samir Hanash, M.D., President and Chair, Human Proteome Initiative Committee, Professor, Department of Pediatrics, University of Michigan

We are now going to shift gears from talking about procedures and approaches that have been tried and tested over many years to talking about the promise of something on the horizon.

We should put the promise of molecular markers for breast cancer or for other cancers into the perspective of what we see them contributing to cancer management in the next 5 or 10 years. Such markers might help us to screen for, and make a diagnosis of, breast cancer by using a blood specimen to make a molecular diagnosis. The same factors, proteins, or molecular elements that are used to make the diagnosis may well help with imaging. If a factor is detectable in the circulation, then presumably it is coming from the tumor itself, and it might be tagged in a way that allows locating the tumor. So there is a continuum from molecular diagnosis in a positive screen, to visualizing tumor cells or locating the tumor, and then to arming the same factors with something that is toxic targeted to those tumor cells. Perhaps, quite a few years down the road, we may not need the imaging component, but at present it is extremely reassuring to be able to localize the tumor after a positive screen before proceeding to any kind of therapy.

How can we come up with novel diagnostics for breast cancer? There are many strategies available, although they have not been implemented to the extent that they should have been, whether because of a lack of funding or lack of resources of other kinds, I'm not sure. I think there is a tremendous opportunity to find the features of breast cancer with the greatest diagnostic promise.

We know a lot about breast tumors and about the genes that are expressed in these cancers. We have an opportunity to look for those genes, and perhaps the proteins, that are expressed in tumors but not in normal tissue, and ask which ones of those could have diagnostic potential, which ones are shed into, and could be detected in, the circulation.

We can also place tumors, or tumor cells, in incubation media to allow them to secrete their products into the media. After identifying those products in the media, we try to detect them in the circulation. And an even more direct approach is to look for the proteins in fluids such as nipple aspirate. With a complete list of those proteins, one can check for their presence and diagnostic potential in the blood.

If the ultimate goal is to be able to take a blood sample and make a diagnosis for breast cancer, why not, as another strategy, simply profile blood proteins and ask which ones can be detected in the blood of subjects with breast cancer and not in controls? Harnessing the immune system could represent yet another strategy. If there are tumor proteins that are aberrantly expressed or are abnormal, then the immune system, which is not compromised early during tumor development, may well react by producing antibodies to those abnormal proteins. If those proteins are identified, they could be put on a chip, and a drop of blood or serum could be reacted with them. A positive reaction would confirm that there are tumor antigens (proteins) present.

These are all logical ways to use current technology to find novel markers for the early diagnosis of breast cancer, although I would say that the entire present effort is really very modest, and a lot more effort has to go into systematically searching for those types of markers that would help us to diagnose breast and other cancers early.

Recall how we started the human genome project 15 years ago. It was with the promise that knowing all the genes in the genome surely would allow us to understand and cure diseases. Of course, now we realize that things are a lot more complex, that while there are perhaps only 30,000 genes, those 30,000 genes produce upwards of a million different proteins. I believe we should now develop strategies that allow us comprehensively to identify and characterize all the proteins being produced by breast cancer cells as a more direct way to find those particular proteins that could be promising diagnostic or therapeutic targets.

We and other groups are doing that type of work. We assume that the proteins that are in the circulation are coming primarily from the cell surface, and

we think that in order to image breast cancer, knowing what is on the cell surface that could be targeted with an imaging agent would be very important. So, why not have a strategy to identify all the proteins expressed on the surface of breast cancer cells? Many technologies are available at the present time to do just that.

One strategy is to tag all the proteins on the cell surface with tagging agents like biotin. Then those cells are broken open and all the biotinated proteins are captured and systematically characterized. We have been working with lung cancer cells. There are thousands of proteins on the surface of these cancer cells. Although not all of them have diagnostic potential, surely those that are uniquely expressed on the surface of such cancer cells could have diagnostic as well as imaging and therapeutic potential.

An initiative to characterize all proteins expressed on the surface of breast cancer cells would have tremendous benefit through identifying those subsets that are important for diagnosis, molecular imaging, or therapy.

There are two or three groups at the present time that are investigating proteins that are secreted by tumor cells. Tumors that have been excised from patients are divided into small pieces, placed in incubation medium, and incubated. They release proteins or other factors which are then collected from the fluid and analyzed through proteomics. A comprehensive mass spectrometry directed approach identifies each protein, and this information is compared with our body of knowledge on gene and protein expression resulting in the selection of proteins that are candidates as diagnostic or therapeutic markers. So, this approach involves identifying the proteins one by one, and, once identified, determining whether they are present in the circulation and characterizing them.

One approach that seems to have attracted attention recently involves profiling the serum of breast cancer patients compared to controls through proteomics technologies. This approach identifies those proteins that are distinctive in the serum of breast cancer patients and, therefore, have the potential for making an early diagnosis. The challenge in this approach is that the marker proteins of interest are mixed in with 6 or 8 very high abundance proteins that make up about 90 percent of serum, medium abundance proteins that make up nine or ten percent of serum, and the many, many low abundance proteins that are found in the final one percent fraction of serum.

From a technology point of view, finding the very low abundance proteins among all of the more abundant ones represents quite a challenge. Figure 2.12 illustrates this problem. There are at least 5,000 proteins that are more abundant than a marker like prostate specific antigen (PSA), which is present in serum in the picomolar range.

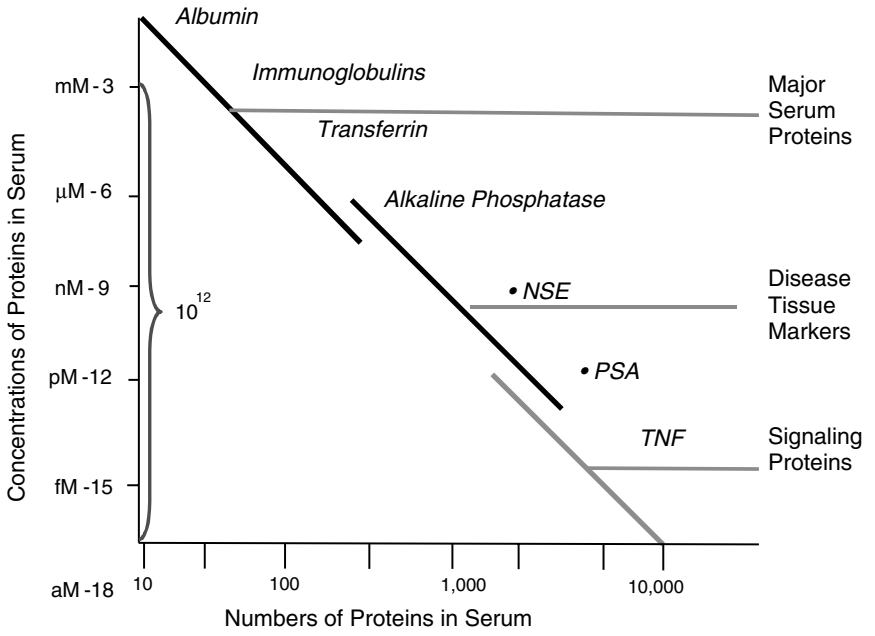

FIGURE 2.12 Disease markers like PSA are present in trillionth molar concentrations as opposed to albumen which is present in greater than thousandth molar concentrations.

A complex and precise technology is required (and fortunately available) that allows profiling thousands of very low abundance proteins and identifying those of high value that may represent markers for different disease states, including breast cancer.

The Human Proteome Initiative, with the support of NIH as well as numerous industry groups, is an effort to utilize all of the proteomics technologies to comprehensively quantify and characterize all the proteins in human serum. This initiative will provide an understanding of the range of variation in our serum and plasma protein constituents. It will develop the knowledge base for the normal serum and plasma proteome and how proteins change with age, with ethnicity, with physiologic states, and with dietary habits, among others. This is a major undertaking which is currently in its pilot phase and has already yielded very interesting findings.

Another proteomics approach provides a comprehensive profiling using multiple tool sets. A tube of serum or plasma is separated based on different

protein characteristics into thousands of fractions, each amounting to about a microliter in volume. Then each one of those fractions is analyzed separately for its protein content, and the proteins are tagged to allow generation of quantitative data. Pre-treatment and post-treatment, or pre-disease and post-disease samples can be tagged with different agents, mixed together, and compared for changes in protein expression. A strategy like this enables sifting through thousands of proteins to pull out the ones that might be associated with a particular condition or disease state such as breast cancer, for example.

With such approaches, we are confident that we can discover the PSA equivalents for breast cancer, for colon cancer and for lung cancer. Specifically for breast cancer, there is an effort now supported by the Entertainment Industry Foundation that is targeting serum profiling using these technologies to find proteins that may be markers for the early diagnosis of breast cancer. This modest effort may provide the proof of principle that this comprehensive profiling of serum enables us to find early diagnostic markers. Furthermore, as I said earlier, this is only one among numerous strategies that could be followed to find potential markers, such as developing a better understanding of the cancer cell itself and what it is expressing on its cell surface, among others. Wouldn't it be exciting if the BCRF and the Entertainment Industry Foundation and perhaps other foundations were to get together to mount a major effort to develop an understanding of breast cancer cells for the purpose of identifying novel markers for early diagnosis and more effective therapy?

And finally, I want to describe more fully another strategy I mentioned earlier and that we and others are exploring. This approach relies on the immune system to tell us whether or not there are tumor cells in a person. Several technologies are available that allow review of all the proteins that could be expressed in a tumor and discovery of which ones are antigenic, that is, cause the formation of antibodies directed against themselves. We can display all the proteins from cancer cell lines on membranes, or blots, and, using sera from different subjects, explore which of the proteins from a particular cancer are recognized by the immune system, that is, act as antigens and generate antibodies.

In one of our studies of breast cancer proteins published three years ago, a particular group of three related proteins, called RS/DJ-1, from a breast cancer cell line, was recognized strongly by sera from four breast cancer patients but not by sera from healthy controls (Le Naour et al., 2001). Figure 2.13 shows the pattern of proteins reacting as antigens with sera from breast cancer patients as spots on the membrane, or blot. Because we detected the presence of antibodies against this protein group from a few breast cancer patients, we asked whether the protein antigen itself was in the circulation of breast cancer patients and could be a potential marker. When we looked for the antigen in the circulation, we discovered that 37 percent of patients with breast cancer had the antigen

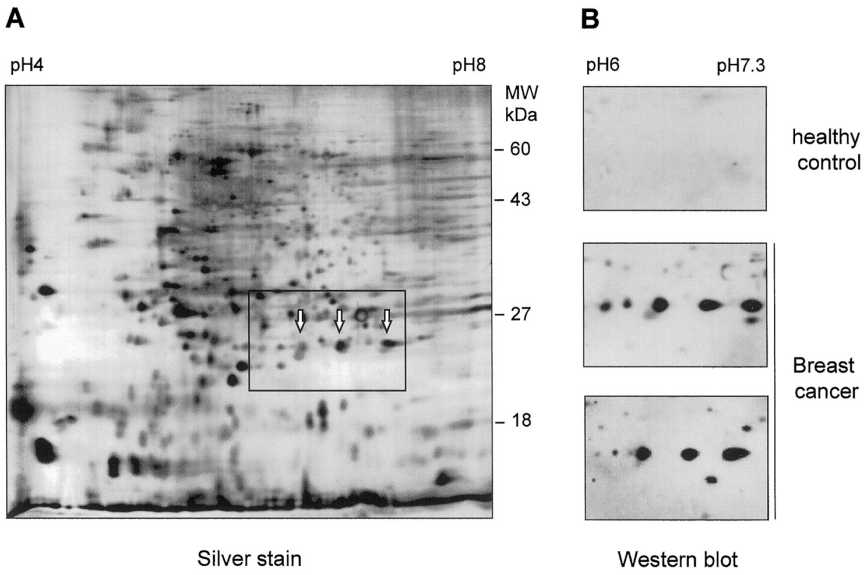

FIGURE 2.13 Western blot showing the presence of protein antigen-antibody reactions as potential markers from patients with breast cancer but not from healthy controls.

detectable in circulation at the time of diagnosis. Some patients had both antigen and antibody, some patients had only antibody, and others had only antigen detectable, and counting those that had either an antigen or an antibody, roughly 50 percent of new breast cancer patients had evidence of this particular molecular marker.

We are very early on in this process of discovering and then validating markers. But this is one among potentially dozens of antigens that could be detectable through antibodies or through the detection of the antigen itself. The goal would be to develop panels of such antigens that together would have the prerequisite sensitivity as well as specificity.

The particular technology that I just illustrated, works very well, but it is tedious, and it is low throughput. To address these problems, we have been able to take the same proteins from tumor cells and rather than put them on membranes dot them on micro-arrays or chips. This allows a more industrialized, higher throughput, higher sensitivity process. Specifically then, we are engaged in increasing the efficiency with which potential markers can be diagnosed.

In the protein micro-array approach, we divide the proteins into several

thousand fractions, and the fractions are arrayed on a chip to produce microarrays, that is, little slides, that contain the entire breast cancer proteome or the colon cancer proteome. For example, our colon cancer chip arrays 4,000 protein fractions originating from colon cancer cells; it can be incubated with serum from a new colon cancer patient. We ask which of the 4,000 fractions the immune system recognizes by an antibody reaction. The answer turns out to be about 50 or so, and we have determined that this is reproducible from array to array. The ultimate goal is to produce chips dotted with protein antigens that can be incubated with one microliter amounts of serum from different subjects, and, based on the pattern of reactivity, can define a molecular signature of colon, breast, lung, or other cancer.

Although this is potentially far reaching, we are extremely early on in this process. We are beginning to learn what is making these proteins detectable in the circulation, and why the immune system is reacting against them. We think that there is tremendous processing of proteins, that a gene does not just encode for a single form of protein, but for something that the cell turns into many different forms. Some of those forms have associations with cancer and can stimulate immune reactions, but we have a lot more research and validation to do. Nevertheless, I am optimistic that we can find markers that are truly useful, better than PSA, for example, and that even as early in the discovery phase as we are today, a targeted approach toward funding the most informative markers for the early diagnosis of breast cancer could help to save women's lives

Bringing New Technologies into Service: Better Quality for Women Through New or Improved Technologies

Sean Tunis, M.D., M. Sci., Chief Medical Officer and Director, Office of Clinical Standards and Quality, Centers for Medicare and Medicaid Services

I am going to give you a framework for thinking about Medicare reimbursement policy for new technologies, focusing on some of the complexity that may not be well known, on some very fundamental statutory and regulatory barriers, and on the legal authority the program has to pay for screening and early detection technology.

Medicare reimbursement falls into five components listed in Box 2.6, each about as complex as the whole. Regulatory approval of payment for technology involves first of all approval by the FDA for at least one use or indication. It need not be the use that Medicare pays for. Payors like Medicare can and do

BOX 2.6
The Five Components
of Medicare Reimbursement

1) Regulartory approval (if applicable)
2) Benefit catergory determination
3) Coverage
4) Coding
5) Payment

routinely (but not always) cover off-label indications, both diagnostic and therapeutic. But if the technology falls under FDA statutory authority, Medicare requires that it be approved for at least one indication. There are ways in which changes in FDA regulatory policy related to technology can influence payment policy. I will discuss this further later, but it is particularly relevant in the case of breast cancer and the coverage of mammography.

The benefit category is the next step, and a major one, in the reimbursement cascade. Medicare is a defined benefits program, meaning the benefit category has to be defined in statute in order for Medicare to pay. Benefit categories include, for example, inpatient treatment, outpatient treatment, or durable medical equipment. And as you know, a prescription drug benefit was added in September 2003. Before the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 (MMA, P.L. 108-173), Medicare could not pay for outpatient prescription drugs, no matter how needed or effective they were.

Diagnostic services are a Medicare benefit category, but screening and preventive services generally require separate legislation. Definition of a service as diagnostic is important because Medicare can pay for diagnosis but not screening. Medicare could pay for breast cancer diagnosis, therefore, including diagnostic mammography, by the use of new, not currently employed technologies.

Screening mammography, however, was added by a change in Medicare law which described screening mammography narrowly and in such a way that the precise definition was left to the public health law and the FDA. That is why Medicare can pay for screening mammograms, but it is also why, as I noted earlier, other technologies besides screening mammography for early detection of breast cancer could not be covered without a statutory or regulatory change at the FDA.

The actual language from Medicare law, section 1861(jj) states that the term screening mammography means a radiologic procedure provided to a woman for the purposes of early detection of breast cancer and includes a physician's interpretation of the result of the procedure. The term radiologic procedure is not

defined in the statute or any Medicare regulations. Instead, it is defined in the Public Health Service Act and FDA regulations pursuant to that Act which limit the term to the standard mammogram, not PET scan, not CT, not MRI, and not ultrasound, among others. There is nothing, then, in the Medicare statute or regulations that would prevent inclusion of a much broader range of imaging technologies under the current statutory authority for paying for screening mammography if the FDA changed the definition of a screening mammogram as embedded in FDA regulations defining a radiologic procedure. Otherwise, it would probably require a statutory change to have any new breast cancer screening techniques paid for beyond the standard mammogram. At least, as I surveyed relevant staff in CMS, this seems to be the correct interpretation.

As I said, the definitions of screening and diagnosis are important in determining Medicare payment. Diagnosis means a test that is done in the presence of signs or symptoms of disease. In other words, if there is an abnormal finding on a mammogram, any technology used to evaluate that abnormal finding is considered diagnostic and a coverable benefit. Medicare may refuse to pay for a diagnostic technology or procedure based on a decision that it is not reasonable and necessary, but at least it is a coverable benefit as opposed to a screening procedure in a healthy woman that discovered the abnormal finding (unless that screening procedure had been added by specific statute).

For example, a very strong family history does not qualify any test as diagnostic. A woman could have the strongest possible family history of cancer and performing what could be considered a diagnostic study in the absence of signs or symptoms of disease or personal history of cancer would be considered screening and would not be coverable by Medicare (again, unless that kind of study had been specifically added to coverage by special statute).

Last year, CMS considered adding testing for diabetes as a benefit for patients with risk factors, but no signs or symptoms of the disease, and explored a regulatory change to achieve this. We discovered we could not add this benefit without a statutory change, and so diabetes screening was included in the MMA.

I suppose that, in theory, Medicare might change the rules to consider interventions to discover disease in a particularly high-risk situation as diagnostic (and reimbursable), but such a rulemaking process would not necessarily be a more efficient, faster process than a legislative change.

Unlike mammography, the statutory benefit for colorectal cancer screening provides that lab-based fecal occult blood testing and other screening tests as determined by the Secretary in consultation with experts are covered. Therefore, if additional new technologies, virtual colonoscopy, for example, met the standard of reasonable and necessary, they would be potentially coverable under the statutory authority for colorectal cancer screening. But, unfortunately, to emphasize what I have said, the mammography screening benefit's statutory language does not allow Medicare to add other breast cancer screening technologies.

I have talked a little of the reasonable and necessary concept, which is the subject of technology assessment and evaluating medical benefits. The Medicare statute provides payment only for things that are reasonable and necessary for diagnosis and treatment of illness and injury. That term is not further defined in any official legal documents. CMS, however, uses a standard definition, which is that there has to be adequate evidence to conclude that the item or service improves net health outcomes experienced by patients (such as improved function, quality of life, morbidity, or mortality), generalizable to the Medicare population, and as good or better than currently covered alternatives.

CMS uses a standard evidence-based-medicine framework, relying on the usual rules of evidence, no different than the U.S. Preventive Services Task Force. There is no formal economic analysis done as part of a reasonable and necessary determination. Although, there is no legal prohibition against considering costs, it is longstanding Medicare practice not to do so when making coverage decisions. A technology or procedure could cost $5,000, $50,000, or $500,000 per life year saved; it would meet the test of reasonable and necessary in any of these instances if it improved health outcomes.

Coverage decisions can be made at the local and/or national level. Many people think that reasonable and necessary reviews at the local level apply a somewhat lower evidence standard of proof and may rely more on expert opinion. The usual view, therefore, is that, in terms of introducing new technologies, bringing those in through the local contractor process is less burdensome than coming in for a national coverage decision.

The national coverage process is diagramed in Figure 2.14. It involves a formally defined series of steps for submitting a request for coverage, so that an evidence review can be referred to an outside advisory committee or a formal technology assessment carried out.

In reasonable and necessary decisions for diagnostic tests (like diagnostic mammography, for example), CMS looks for studies of test performance, classic sensitivity and specificity, for impact on patient management and outcomes, which may depend on whether or not there is a beneficial intervention available, or asks if the information itself provides a benefit. Certainty in diagnosis may be useful by itself, or information about diagnosis or prognosis may influence decisions about the use of other health care services, institutionalization, or other care. If there is empirical evidence that knowledge of diagnosis influences the care or quality-of-life of the patient, that certainly would be a potential basis for meeting the reasonable and necessary standard.

In some situations, however, like PET scanning for Alzheimer's disease, where treatments are relatively ineffective and not particularly toxic, it is preferable simply to go ahead and treat based on the clinical story rather than run the risk of withholding treatment based on a possible false negative scan. In this situation, therefore, the test does not actually improve a health outcome.

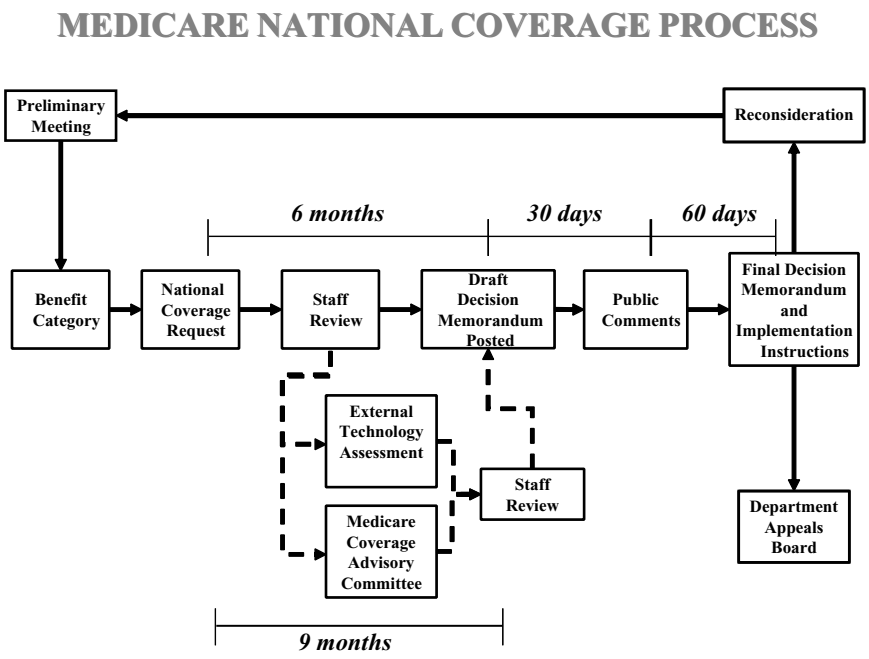

FIGURE 2.14 The national process for determining Medicare payment coverage.

Sometimes, a CMS review of coverage results in a split decision, such as the coverage decision on PET for breast cancer. It is covered for diagnosis not for screening, consistent with the usual rules on benefit categories I discussed earlier, but it is covered as an adjunct to standard staging in loco-regional or distant recurrence and monitoring for response to therapy. It is not covered for evaluating abnormal mammograms or palpable breast masses or for evaluation of axillary lymph nodes to decide on lymph node dissection.

Assuming that CMS could make reasonable and necessary decisions about screening and early detection of breast cancer, we still need to know who is doing the evidentiary studies, what is the quality of the studies, and what methods are required to study comparatively sequences of multiple studies. It appears that handling these questions is not part of anybody's research agenda. As the IOM report points out, the funding for applied clinical research, how to use technologies once they are developed, is really nobody's domain. This is what I am calling the systematic gap in evidence from applied clinical research.

Decision makers are interested in using high quality evidence to support clinical and health policy choices, but the quality of available evidence is inadequate (Tunis et al., 2003). The whole concept of practical or pragmatic clinical

trials is one that has been talked about since the original discussion of pragmatic attitudes and therapeutic trials in 1967 (Schwartz and Lellouch, 1967). I am talking about the gap in studies in which the hypothesis and design are developed specifically to answer the questions faced by decision makers. These trials select clinically relevant alternative interventions to compare, include a diverse study population with broad patient eligibility recruited from heterogeneous "real world" practice settings, allow natural variation and are minimally intrusive on care, and collect data on a broad range of health outcomes both functional and economic (Tunis et al., 2003). I emphasize that the key is clinical research that is designed to answer the questions of decision makers like patients, like clinicians, and like payors and purchasers. That research looks very different from clinical research that is designed to answer fundamental questions about etiology, causation, and the like.

CMS, under Dr. McClellan, is very interested, given that we are a major consumer of information needed to make decisions about payment and coverage, in trying to find ways to get into this business. We are working in a number of ways to try to facilitate the support and infrastructure for practical clinical research, research on comparative effectiveness, through Section 1013 of the MMA. We have new collaborations with NCI, NHLBI, and FDA to focus on this issue, looking into alternatives or modifications to clinical trials, registries, quasi-experimental studies, among others, to try to improve the evidence base. We also have a lot of interest in exploring coverage under protocol, that is, paying for emerging and promising technologies, but only in the context of a clinical trial or some kind of clinical study with systematic collection of evidence.

I would like to conclude with some general comments on health care costs. We all know that we are spending a lot of money on health care, so it is important that we know the risks and benefits of the interventions we are using. While I said there is no explicit consideration of costs in Medicare coverage and payment, cost is the universal context of health care decision making. Both in the IOM report as well as in any other discussion, we must think about the economic implications of new or added technologies, the additional clinical value or the additional social value, and how that is related to the investment. There are many other unmet needs in society for health care services that should be balanced against additional spending for improvements in breast cancer technology or increases in payment to improve the quality of mammography.

DR. PENHOET: We have time now for questions and answers from our last three speakers.

DR. DUNNICK: What is the timing regarding the suggestion about testing technologies in practical clinical trials since we cannot test these with the classical randomized controlled trials or we will be waiting 10 or 15 years for results?

DR. TUNIS: One of the limitations of the evidence-based framework for coverage and payment policy is the long time it can take to prove something actually works, time during which introduction into practice should have occurred. The answer to getting relatively quickly from proof of principle at the bench to early experience at the bedside is to have some framework, as I have described, where there is a subset of technologies that look promising and could be reimbursed in the context of defined protocols in order to find out which actually work and which do not. So often at present as technologies are first introduced into practice, there is uncoordinated trial and error that produces no information about whether they are or value, and that wastes a lot of money and time.

DR. ESSERMAN: Our report states that the organization of screening mammography and breast cancer care is important and could make for greater cost-effectiveness and better quality. But there is no funding or infrastructure to support that. I'd like to hear you comment on that.

DR. TUNIS: I think we are actually at an inflection point in terms of payors thinking more about how payment policy might promote the efficient organization of care. Historically, Medicare has been a resource-based payor, that is, paying for resource consumption. In that context, there is no place for discussing how to pay for delivery of a service in a high quality way. Now, however, there may be an opportunity to make a case to both private payers and CMS for payment for models of care that are efficient. We can look at our regulatory and statutory authority and find out how we might facilitate that. Medicare now has a new Section 721 which allows a new payment mechanism for coordinated care for chronic illness, moving away from resource consumption. If you can do it better and cheaper, we will find a way to financially reward that.

DR. ESSERMAN: In the committee, we considered whether proteomics would have to go head-to-head with mammography, that is, would have to have equivalent sensitivity and specificity. But that could hamper development of inexpensive and easy tests that have very high sensitivity and low specificity for use in a primary-secondary screen system. That is partly an FDA issue and partly a CMS issue; but what would enable or support this kind of integrated approach to combining tests and looking at more clever ways of harnessing technology?

DR. TUNIS: The standard framework for evaluating new technologies does not easily apply to such staging of tests. I don't have the answer to how to do that. If you came up with a framework by which that kind of question could be answered, we payors would have to look at it, but at the moment we have rather simple-minded evidence standards that do not apply very well.

DR. SMITH: Dr. Fletcher, I thought you addressed the issue of risk stratification very nicely in terms of the Gail model in context and its application to identify a population for study. The problem is, short term risk is deceptively

small, long term risk is deceptively large, and that makes it very hard for women to get a handle on how to think about risk mathematically for a disease that means a lot to them.

As risk grows over time, screening needs to be thought of in terms of social insurance. Most women will not develop breast cancer in their lifetime, but screening can measurably reduce the risk of being diagnosed with advanced breast cancer that could result in a premature death. With respect to all-cause mortality, ultimately only about three and a half percent of women die of breast cancer, but that is not nearly as important as the contribution that breast cancer makes to premature mortality. In any given year, deaths of women diagnosed with breast cancer in their forties account for about 16 or 17 percent of all breast cancer deaths.

We looked at the two-county trial to estimate the effect of a reduction of risk of dying of breast cancer on all-cause mortality. As you said, women in their forties have a very low risk of dying of anything. But among the causes of death breast cancer is quite a significant factor. The breast cancer mortality reduction we saw in that group meant a reduction in risk of dying of anything by about 50 percent.

I am wondering how you would reconcile the near-term risk versus the long-term risk against the background of something simple to do that can ensure, even though you have a low probability of dying of anything, that you have significantly reduced your risk of dying of breast cancer by participating in screening.

DR. FLETCHER: I think you are illustrating why this is such a complicated, almost counterintuitive area. The prevention paradox is that for the vast majority of people undertaking an intervention, for example, screening for breast cancer, there is no benefit. Yet, for the women who do benefit, it may be quite substantial.

As you said, in the younger age groups, we are talking about 10 to 15 percent of deaths caused by breast cancer. I think what I showed reaches the same conclusion as what you said. Stratification of risk is going to be tougher than at least I had previously thought. We need groups of risks that are really quite a bit more substantial than most of the risks we have so far identified. Furthermore, regardless of what we come up with in terms of a new model incorporating brand-new technologies, we must validate it not only in terms of calibration, but in terms of the ability to discriminate among women if it is to be used for risk stratification.

DR. PETITTI, Kaiser Permanente: The promise of the ultra-low breast cancer risk group is demonstrated by the study of Cummings and colleagues showing that at some age there is a serum marker (serum estradiol level), which is not based on proteomics, that identifies a group of women that have a very low risk of developing breast cancer over some reasonable time frame, say 4

years (Cummings et al., 2002). When you think about it, the ability to decide when you have the risk of a 70-year-old man and don't need to have mammography is very important. There are analogies in other fields of cancer screening; we are now finding that a 55-year-old woman who is human papilloma virus negative might not need a Pap smear every year, and there are analogies from the cardiovascular field, where someone who has a low density lipoprotein of 80 and a high density lipoprotein of 100 probably would not be a candidate for a screening test for early cardiovascular disease. So I think the ultra-low risk group is as important as the high risk group.

DR. WARRICK: Are low risk, high perception women disproportionately utilizing mammography capacity, and if so, does this explain 60 percent of women reporting having had mammograms to the Behavioral Risk Factor Surveillance System, and only a little more than 30 percent of Medicare eligible beneficiaries getting screened. Should the Breast and Cervical Cancer Early Detection Program be modified based on these findings?

DR. FLETCHER: As mammography utilization increased in this country, for a long time women in their 40s utilized it more than women over 50. I think now, the women over 50 have a slightly higher percentage utilization. And the women over 65 do not utilize it nearly as much.

DR. DUNNICK: Women at a higher risk do have mammography more frequently, but still, 30 percent of them do not have periodic mammography. This difference is present from 41 to 49, but not over 50, which is interesting.

I wanted to ask Dr. Hanash about the exciting material he presented; when can we expect a fusion between basic bench science and a clinically usable test, given the problems of looking at probes for breast cancer, and also the problems that we see with PET scanning and with BRCA positive women.

DR. HANASH: Aside from funding problems, there is a major issue with respect to validation of interesting markers. The initial discovery work is usually done with very contrasting groups, those with overt disease and those who are completely normal, and in the disease group something promising shows up. But if our goal is early detection, the disease group is not representative of an early detection population. Having access to samples from an appropriate early detection population could be highly informative, but those types of samples are scarce and access to them is limited.

For example, we found a few potential lung cancer markers, and asked if we could have access to samples taken over a period of time from subjects who later were diagnosed with lung cancer to see if our markers worked to identify early disease. We were asked about evidence that our markers were effective for early detection. We answered that we did not have such evidence; we hoped testing the samples would provide the evidence. The people controlling the samples would not allow them to be used without some data on our markers' effectiveness in early detection. So, although we ultimately did get some access, initially

we were in a catch-22, having to do our discovery work with samples from not the most appropriate study subjects.

For prospective validation of markers for screening, obviously it is impractical to embark on a validation study in a very low risk population. So there, I think you would have to consider strategies demonstrating that a marker builds on existing tests by improving sensitivity and specificity. It would expedite things if you could demonstrate early on that your panel of markers improved specificity and sensitivity of CT based screening for lung cancer or mammography based screening.

DR. FLETCHER: Randomized trials were mentioned in validating new technologies which, especially for prevention, take decades to complete. But relatively simple evaluations that do not take so long and do not require randomized trials can be carried out for new screening technologies. For example, for any new screening test, it is important to determine the test characteristics, sensitivity and specificity. Sometimes, even if sensitivity and specificity are determined for a new test, the evaluation is carried out in a diagnostic situation on patients who are symptomatic and/or have an abnormality. Assuming that the results of such an evaluation would generalize to a screening situation is dangerous. It may be reasonable to start out evaluating a new test in a diagnostic situation, in which you quickly know those who have cancer and those who do not, to learn how the test performs. But then the test's accuracy must be evaluated in a screening situation, because the spectrum of cancers is likely to be very different in that group. Too often, this kind of evaluation is not being done in a systematic way with newer screening technologies. I do not want us to think that we cannot know anything without a randomized trial. Systematic evaluation of the accuracy of a screening test does not require a randomized trial.

Finally, I just want to remind everybody that sometimes randomized trials end up with rather unexpected results. The Women's Health Initiative (WHI) leaps to mind. Here we thought women my age were supposed to be on long-term hormone replacement therapy to prevent several important chronic diseases, and all of a sudden not only the WHI but the Heart and Estrogen/Progestin Replacement Study and the Million Women Study are giving the lie to that conclusion. I was on the Board of Scientific Advisors at NCI, and there was concern about the high cost of the WHI. In retrospect, the cost was nothing compared to the billions of dollars being spent every year by women for a prevention therapy that we now see in an entirely different light. This teaches us to persist with randomized trials every once in awhile, even if they are expensive.

DR. ESSERMAN: I wonder whether there is a benefit to having a lot of these test sets public, whether something like the NCI's Early Detection Research Network is going to facilitate making the data available not just to the

researchers, but to the scientific community in general? Will that approach to keeping track of and sharing data early on accelerate discovery and deployment?

DR. HANASH: There is this notion that one gene, one protein, one marker may not be enough, hundreds of them together might be required to be informative. If this is the case, obviously having all of the data in the public domain would be extremely useful so others could mine the same set of data using different kinds of software tools and different statistical approaches.

Others have thought, that patterns could emerge that are not understood, and that they will represent the diagnosis for various cancers. I hesitate to recommend that we rely on patterns that are not understood for a diagnosis. I think the resources and technology are available to decipher the unknown features and patterns, to link back to the disease process, so that if it does not look very plausible early on, we know there is a problem.

DR. NORTON, Memorial Sloan-Kettering Cancer Center: I would second that. Whole books have been written about how misleading blind looking at patterns can be and how not understanding the mechanism can lead you far astray in applications of technology developed in one area to other areas.

I want to emphasize that one of the important things about IOM reports is that they can influence policy makers. We have examples in this report of recommendations that I think should be publicized and acted on. One of them is the notion that breast cancer screening, when applied in other countries in an organized fashion, has clearly been shown to reduce mortality. The British epidemiologist, Richard Peto, has shown elegantly over time that as you introduce existing technology and do it in the proper fashion, you see a reduction in mortality. You deal with a country like ours, where screening is not well reimbursed, you see the response. It is a regulatory issue and a statutory issue. The fact is, we know that people are dying because of the mis-application or the lack of application of the technology. It conceptualizes a very important area that we have to address, which is, how do we effect societal change, which means governmental change as well.

I also think that probably the data already exist to answer many of our questions, or at least, the samples already exist. But we haven't heard about HIPAA, the Health Insurance Portability and Accountability Act, which prevents us from doing a lot of the retrospective work that is necessary, to correlate data from samples in serum banks with outcomes to try to address some of the important questions. There, too, we have a barrier that is getting between us and the ability to solve problems

DR. FLETCHER: From the perspective of an epidemiologist, we are running more and more into trouble with HIPAA regulations, too. I certainly hope the research community is going to be able to work to correct some of those.

DR. HANASH: The research community is vested in this, so it has to come from a third party, as opposed to we researchers trying to make a plea with the regulatory agency. The consumer and the public have to participate.

3

Simultaneous Group Discussions with Invited Speakers

First Group Discussion: Delivering Better Breast Cancer Screening Services

Etta Pisano, M.D., Professor of Radiology and Biomedical Engineering, Chief of Breast Imaging and Director, UNC Biomedical Research Imaging Center and Member, Committee on Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis

MODERATOR AND RAPPORTEUR: The first speaker on this afternoon's panel is Stephen Taplin, a senior scientist in the applied research program at the National Cancer Institute, who will tell us about the organization of breast cancer screening services.

STEPHEN TAPLIN, M.D., Senior Scientist, Applied Research Program, NCI: Today we want to talk about how we can improve breast cancer screening by organizing care. I underline that screening is a process that leads to outcomes, not a test. Figure 3.1 presents the steps in the screening process. There are at least four different steps in the screening process: risk assessment, looking at who we are trying to reach; detection, finding an abnormality, where today's focus may be; diagnosis, evaluating the abnormality to find the cancer; and treatment. The transitions between these steps need to be organized as well. Focusing only on improving the steps, not on how women get from one step to another, will not result in improved breast cancer screening.

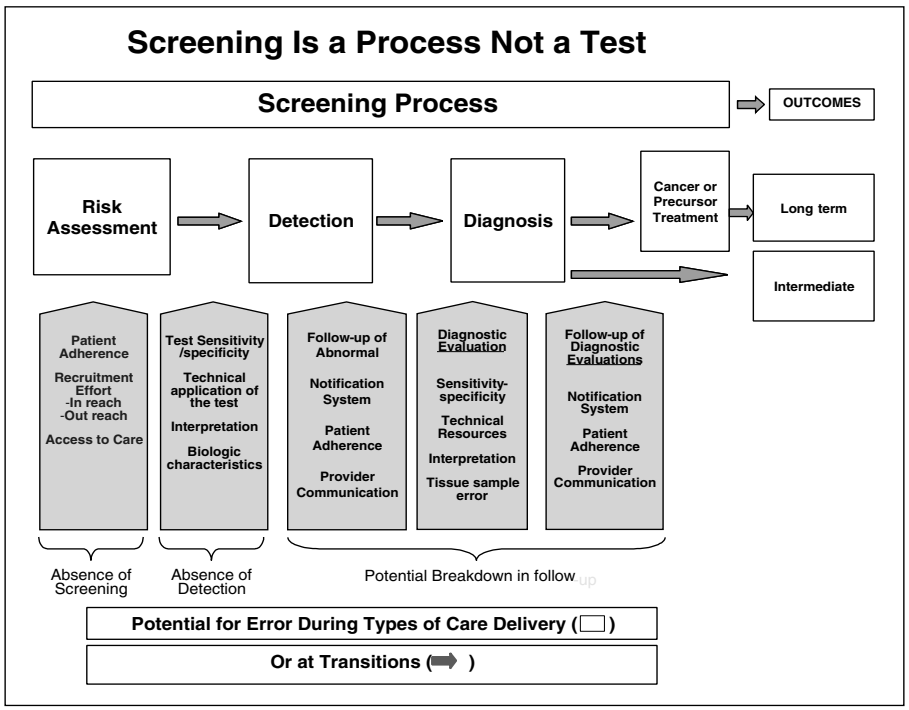

FIGURE 3.1 The screening process.

The whole process leads to at least two outcomes that can be examined - the long-term outcome of mortality and some short-term outcomes, like reductions in late stage disease.

To dissect this process, we can start by looking at recruitment. What are the ways to get people from the population at risk into screening? The next step is the detection process, and for this step we need to evaluate sensitivity and specificity, the quality and validity of the test itself. The third part is follow-up. In a study we just completed, we tried to simplify this process and isolate the problems (Taplin et al., 2004a). We decided to identify the source of all the late stage cases within an organized system. Were they people who were not being recruited, were not being detected, or had a breakdown in the follow-up of their care after a positive screen? Box 3.1 illustrates the sources of advanced cancers in populations with health insurance coverage where 70 to 80 percent of the women reported they had been screened. We found that 52 percent of the advanced cancers were recruitment failures. This teaches us that organized screening must include organized recruitment.

BOX 3.1
Sources of Late Stage Cancers

Failures in the process are associated with poor outcomes:
- 1,347 late stage cancers from 10 integrated health plans.
- Absence of screening (1-36 months)—52 percent of late stage breast cancers.
- Absence of detection—40 percent of late-stage breast cancers.
- Breakdown during follow-up—8 percent inadequate/other follow-up.

SOURCE: Taplin et al. 2004a.

The limitations of mammography are not trivial, however. Forty percent of these women had a mammogram within the prior three years. So reduction of these detection failures by improving the quality of technology is important, although not the whole story. Failure of follow-up, where I thought the action would be, was, in fact, the cause of only about 8 percent of advanced breast cancers in this population.

We need to think about how to change the system. The Institute Of Medicine (IOM) report recommends organized screening as a way to make screening happen in our populations. This builds on several previous IOM reports, beginning with *Crossing the Quality Chasm*, which stressed the need for systematic change to improve quality (Institute of Medicine, 2001). Organizing screening is a dramatic system change.

Then, measurement is critically important to clinicians making changes. Showing people they are making progress, that they are reaching entire populations, is a critical feedback loop. It tells people they are getting results from their actions in a way they otherwise might not appreciate.

European models of care have been mentioned. They demonstrate that organized care does have a definition. Box 3.2 lists that it is about an explicit

BOX 3.2
Definitions of European Organized Screening

European Models of Organized Care (IARC, 2002):
- An explicit policy, with specified age categories, method and interval for screening.
- A defined target population.
- A management team responsible for implementation.
- A health care team responsible for care and clinical decision.
- A quality assurance structure.
- A method of identifying cancer occurrence in the target population.

TABLE 3.1 Results of Organized Screening Programs

Start	Location	Evaluation Design/ Comparison	Outcome	Age	Effect	Confidence Interval
1974	Sweden (Gavleborg)	Geographic comparison	Excess breast cancer (bc) mortality	40-64	.84	(.71-1.00)
1975	Netherlands	Geographic comparison	BC mortality	50-79	.84	(.61-1.17)
1978-1979	Sweden (seven counties)	Geographic comparison	BC mortality	40-69	.68	(.60-.77)
1987	Netherlands	Before/after comparison	BC mortality	45-69	.76	(.53-1.09)
1988	U.K.	Modeled estimate	0/E	55-69	.79	n/a
1990	Italy	Cohort analysis	Mortality ratio	50-69	.75	(.54-1.04)

NOTE: Effect = breast cancer with organized screening/control population; Confidence Interval = 95 percent confidence interval.
SOURCE: IARC, 2002.

policy, specific age categories, a method to do it, and an interval for screening. It is a policy with a very explicit approach and a defined target population, a population at risk as specified in Figure 3.1. It is also about somebody being responsible, a leadership team. These things do not just happen. A quality assurance structure and measures to create feedback are also essential.

So organized screening is happening in Europe. Are these programs having an impact? Of course they are. They have shown a clear trend towards reduce tions in mortality in the populations as illustrated from 1974 through 1990 in Table 3.1.

Can it be done in the United States? I just completed 20 years working at Group Health Cooperative, where we put this kind of program in place, and we are also beginning a pilot project in 20 Bureau of Primary Health Care clinics. Group Health is somewhat unique. However, there are a number of plans around the country that have an organized insurance structure, and, in any event, I think our plan's success was due more to leadership and paying attention and commitment to a direction than to the structure of our medical system.

We organized five mammography screening facilities within our system which serves 400,000 people in the Northwest with more than 70,000 women age 40 and above. About 35,000 women were screened each year in that population. We created a multidisciplinary team and a team for leadership. There were different providers involved in each of the steps. When we started to influence, and get feedback on, what was happening in our population, it was not a big

problem to get these people together to ask what we were doing, what we needed to do.

We had a group that included surgeons, oncologists, nurses, radiologists, administraters, and primary care physicians, all working together to organize this care. We had groups in each region delivering the care. There was clinical leadership at a facility which involved the radiologist and primary care physicians as well as the nursing staff. And we had an information system which mailed reminders. The critical part is identifying the population and communicating with the women. We organized outreach to look across the whole process and explore how to improve it.

One of the first things we learned was that follow-up was not coordinated, and there were people falling through the cracks. So we created a system in which there were nurses in the radiology center who were responsible for the follow-up of all positives. All positive mammograms were in a database, and the nurses took responsibility for communicating with the primary care physician and the surgeon.

We carried out a number of studies funded by the National Cancer Institute (NCI) to look at recruitment. A reminder postcard was found to be effective in improving recruitment (Taplin et al., 1994). We did a risk survey. Collecting risk factor information from a survey and informing women about their personalized risk increased the likelihood that they would come in for mammography to 66.7 percent compared to 42.9 percent among controls, who received generalized risk information (Curry et al., 1993). We found that a simple call, in which the woman was asked to come in and was scheduled, was as effective as a call addressing all the care issues. The reminding phone call itself was sufficient (Taplin et al., 2000).

Then we turned to detection. We measured sensitivity, specificity, recall rates, and positive predictive values, and we provided yearly reports back to the radiologists including all the false negatives, which is more than is required by the Mammography Quality Standards Act (MQSA). Radiologists then went through their own quality assessment program. We looked at a method of improving clinical image quality and reported that interval cancers were more likely to occur in mammograms of poor quality (Taplin et al., 2002), and we are currently studying computer assisted detection. Then we conducted a teaching session with our technologists to try to improve clinical image quality.

We evaluated follow-up and treatment. As I said earlier, our nurses assumed responsibility for communicating both with the primary care physicians and the surgeons to ensure follow up. If it was not occurring, they contacted the women. Group Health has also been looking at how treatment is organized. We already know that within our group the odds of breast conserving therapy are about 300 percent higher than in the surrounding community.

Over a 17-year period, we systematically changed the entire system for individuals, by surveying and by recruiting; for the physicians, by giving them feedback about patient participation and results of screening; for the organization as a whole, by a steering committee and a multidisciplinary team; and by creating an information system.

Did it have an impact? Absolutely. Our screening rates (ever had a mammogram) among women age 50 and above increased from under 50 percent to over 80 percent between 1986 and 1990, and we had similar increases in rates for women between 40 and 50 years of age and for mammograms within the last two years (increased from about 26 to 51 percent in women age 50 and above). We also reduced the numbers of women with late stage disease as shown in Figure 3.2. Our report of these data provides evidence that enrollment in an organized screening program is associated with increased likelihood of mammography and reduced odds of late-stage breast cancer compared to community controls (Taplin et al., 2004).

We can't describe the total use of resources in the surrounding community, because we don't have individual level data, but my suspicion is that our program consumed fewer resources, that is, was more efficient. So organized care

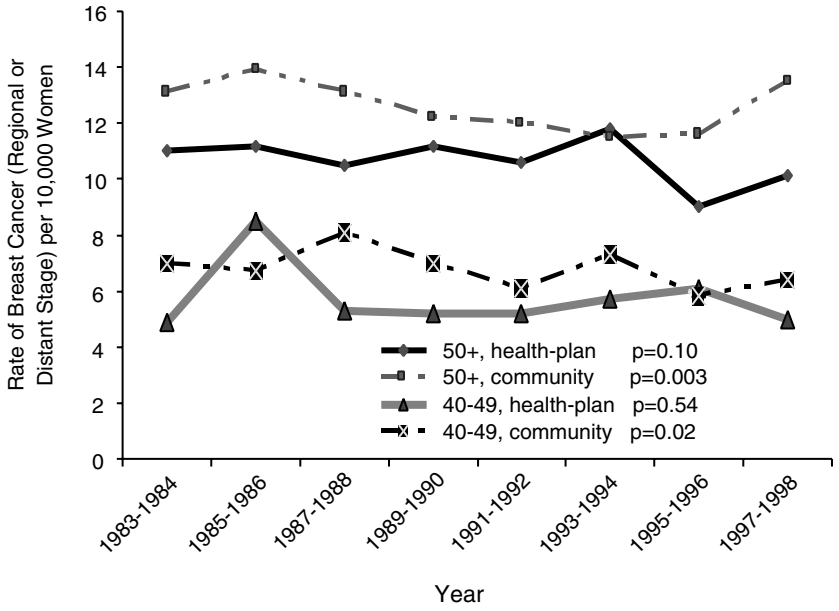

FIGURE 3.2 Group health rates of women found with late stage disease are lower as a result of early detection.

has advantages within this health plan in promoting efficient use of screening, affecting screening and screening interval rates, and in reducing late stage disease.

Having made the case in one place, what is the next step? Is it possible to organize screening in other settings? We turned to the cancer collaborative, which is a consortium of agencies, the Bureau of Primary Health Care, the NCI, the CDC, and the Institute for Health Care Improvement which have joined in a 2-year effort to change screening in a comparable way within the Bureau of Primary Health Care.

The Bureau's clinic sites are intentionally spread around the country. Of the 800 clinics, there are 20 sites participating, so it is a small proportion, but it is the pilot proportion. The 800 clinics as a whole serve more than a million women over age 45, so there is a chance to have an impact on a large population of people. The collaborative's program makes the policy explicit, sets targets for improvement, and then asks about measurement. We create a leadership and implementation team within the primary care group. That team is the physician, the nurse, the PA, the medical records person, and the receptionist. We organize the recruitment, the follow-up, and the referral for treatment. Then we encourage regular changes within the clinic in order to achieve these, and we create a data system to identify how many people receive care. We emphasize systematic reorganization and practice teams. We meet monthly with these teams. We have three sessions within the year in which we discuss progress and assess what they are doing. We then meet on a monthly basis as they put the new plan into place. The bottom line is they are being asked to look at what they are doing, change it, and measure what the results are.

So in conclusion, we can systematically change the screening process. It does not require rewriting legislation; it requires the will and the leadership to do it; it takes time, and it takes data. We need also to address barriers to collaboration and whether we can create an environment in which quality improvement is encouraged and reinforced. Those are important questions for us and for our society.

REBECCA SMITH-BINDMAN, M.D., Associate Professor, Radiology, Epidemiology, and Biostatistics, Obstetrics, Gynecology, and Reproductive Medicine, University of California, San Francisco: There has been an enormous amount published over the last few decades about who is getting screening mammography. There have been several predictors of screening noted—age, race, ethnicity, having a usual source of care, rural residence, as well as financial barriers. I think there is a general belief now, however, that the differences by these predictors have declined as a result of numerous mammography outreach efforts.

Our knowledge of the current status of mammography is based largely on two very widely cited surveys, the CDC's Behavioral Risk Factor Surveillance

System and National Health Interview Survey, both of which assess mammography use annually. The surveys have found that the historical discrepancies in mammography use that have been seen by race and ethnicity have declined significantly. According to these self-report surveys, most eligible women are now getting mammography; 70 to 80 percent of women report that they have had a mammogram in the last two years, and there is no reported difference by race and ethnicity. In fact, some minority groups report higher rates of screening than white women. So many people, including policy makers, have concluded that compliance with recommendations for screening mammography is no longer a problem.

We have not discussed differences in cancer statistics a lot today, although Dr. Esserman touched on this in her presentation. However, I think it is understood that there are substantial differences in breast cancer outcomes and breast cancer detection rates by race and ethnicity. In general, non-white women have more advanced disease at diagnosis. There have been improvements in breast cancer mortality over the last decade, but SEER statistics show that the improvements have been largely limited to white women. Mortality curves have been essentially flat for other racial and ethnic groups.

It seemed to me that one might question the value of mammography if mammography use is now the same (and high) among different racial and ethnic groups, but differences persist in breast cancer mortality and tumor stage at diagnosis. Clearly, if mammography were working, we would expect breast cancer mortality rates to decline coincident with improvements in screening mammography rates.

One possible explanation for the failure of mortality rates to decline among racial minorities along with their higher use of mammography is that estimates of their use of mammography based on self-reports may be inaccurate. I think there is growing concern that this might be the case, and that women, particularly minority women, may overestimate their use of mammography.

So, I have been interested in investigating whether there are persistent differences in the use of mammography. We have just completed a study that evaluates screening mammography use among a large number of racially and ethnically diverse women diagnosed with cancer. This study examines recorded mammography use from medical records, and, therefore, we believe these data are more accurate than self-report data. We used data from the NCI funded Breast Cancer Surveillance Consortium which comprises mammography registries in seven states and is probably the largest data set available to assess actual mammography in the United States.

In this data set, we learned about mammography use based on medical records, radiologist reports, and a survey that each patient completed every time she had a mammogram (such as a patient self-reported breast mass at the time of

a mammogram). These data allowed us to assess mammography use in a more detailed way compared with the self-report surveys that are relatively crude in terms of their assessment of mammography—that don't differentiate between a screening mammogram, or a diagnostic mammogram, or whether there was a mass at the time of mammography. Clearly if a woman has a mass at the time of a mammogram, it should not be considered a screening test.

Cancer outcomes are complete in this data set, since over 95 percent of cancers are ascertained based on linkage to cancer registries. The data describe approximately 900,000 women who are racially and ethnically diverse. In these women ages 40 to 85, there were approximately 26,000 breast cancers diagnosed.

Table 3.2 displays the characteristics of tumors in these women by race and ethnicity. These numbers should be, and are, very similar to recent reports using SEER data. They display the adjusted odds of advanced stage cancer by race and ethnicity using white women (set at one) as the reference. Essentially, African American, Hispanic, and Native American women are at increased risk of having an advanced cancer at diagnosis, and such tumors are less likely to be curable. These are the results you would expect, looking at current population tumor registry data.

We next looked at the mammography use among these women in the five years prior to breast cancer diagnosis. We categorized women into five groups based on their screening frequency. The most screened women had a mammogram a year before cancer diagnosis. The least screened woman had not had a screening mammogram for at least five years prior to cancer diagnosis. We considered women to have inadequate mammography if they had either never had a mammogram, had not had a mammogram for at least three and a half years, or had their first mammogram after age 55, or only coincident with the diagnosis of cancer.

TABLE 3.2 Breast Cancer Characteristics by Race and Ethnicity with White Women as the Reference (adjusted odds ratios)

	Large 15mm	Advanced Stage	High Grade	Lymph Node +	Symptomatic
White	1	1	1	1	1
African American	1.45	1.60	1.80	1.25	1.18
Hispanic	1.40	1.44	1.20	1.21	1.26
Native American	1.47	1.22	1.60	1.02	1.90
Asian	1.04	1.03	1.31	.86	1.09

There were substantial differences in mammography use by race and ethnicity. Compared to white women, all minority women were less likely to be screened regularly and more likely to be screened infrequently. In comparison to whites, the odds ratios varied from 1.4 to 1.8, and these ratios suggest minority women were around 40 to 80 percent more likely to not be screened.

We then looked at tumor characteristics after adjusting for mammography. We asked whether women who were similarly screened would have similar types of cancer, or are minority women who are similarly screened still at increased risk for advanced cancers. The latter would suggest a biological explanation, the former discrepancies in mammography use.

Once we stratified by mammography, the differences in tumor size, stage, lymph node involvement, and symptoms by race and ethnicity were reduced or eliminated. Thus no matter what a woman's race or ethnicity, similarly screened women had similar types of tumors. Interestingly differences in tumor grade persisted even after adjusting for mammography.

Figure 3.3 shows the percent of women with large tumors in each racial and ethnic group by use of mammography. From left to right, the groups go from more to least use of mammography. The percentage of women with large tumors increases as mammography decreases, and large tumors were found in about 80 percent of weomen who were never screened.

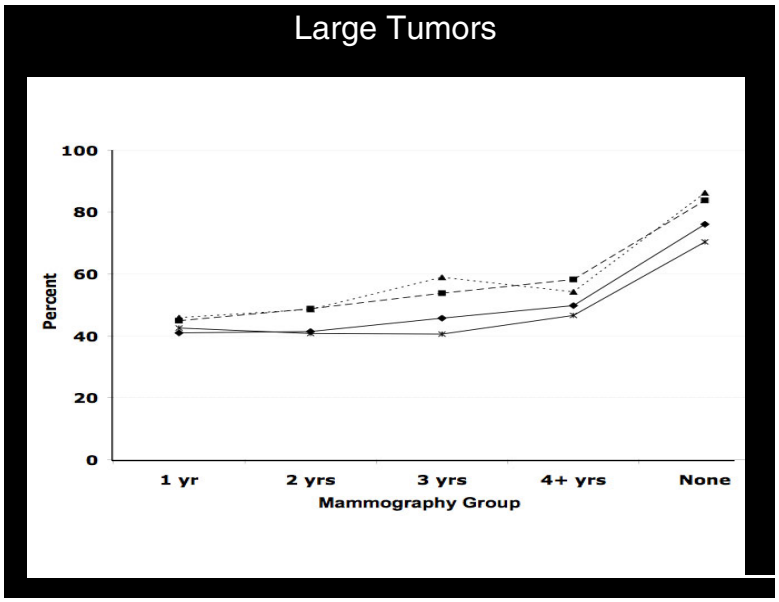

FIGURE 3.3 Increasing tumor size with increasing interval since mammography.

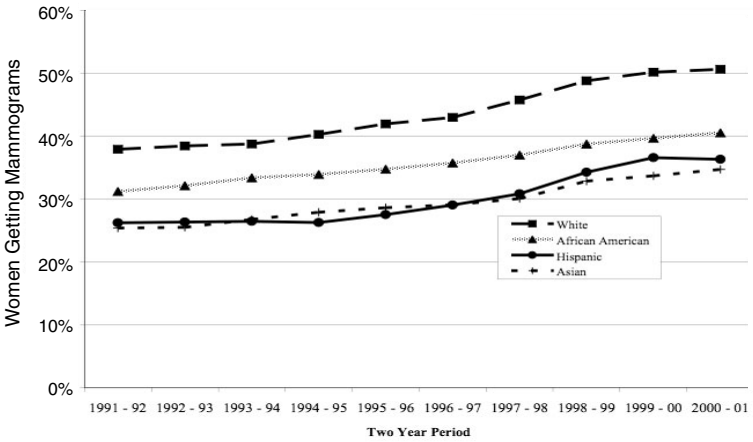

FIGURE 3.4 Racial and ethnic differences in mammography in Medicare beneficiaries.

However, once plotted by mammography, there are few differences by race or ethnicity at each screening level. Similar results were found for advanced stage, lymph node involvement, and symptoms. Tumor grade did not change much from the most recent to the most distant mammography groups, and African Americans persistently had higher grade tumors as noted earlier by Dr. Esserman, apparently for biological reasons. It is clear that mammography use is associated with size of tumors, but differences by race and ethnicity, represented by the four different lines of the figure, are no longer significant.

In summary, there are persistent and dramatic differences in cancer characteristics which are reduced or eliminated when data are adjusted for mammography use. Therefore, mammography appears to be in large part causal for the differences in tumor characteristics by race and ethnicity. Since mammography clearly contributes to racial and ethnic differences in mortality and to reduction of mortality in general, it is vital to increase the regular use of screening. Having had a mammogram once does not protect against advanced stage cancer, regular screening is required. We consider a 3-year interval to be a minimum requirement.

I now turn to mammography use among elderly women. Medicare billing records are a great source of data to assess population-based use of mammography in Medicare eligible elderly women. Clearly mammography rates in this population have increased over time. However, the overall rates are substantially lower than suggested by self-report surveys. In contrast to self-report data, data from Medicare billing records suggest substantial and persistent racial and ethnic disparities.

Figure 3.4 illustrates biennial mammography use over time in women aged 65 to 69 in whom the evidence is clearest that mammography is beneficial. Rates of mammography have increased, but even as recently as 2001, those rates are only 50 percent, far lower than the 70 to 80 percent from the self-report surveys. There is obviously a significant gap between white and African American women which is growing rather than shrinking.

In summary, there are persistent disparities in who is getting mammography. Repeat mammography is very important, and it is less frequent than widely believed. Dr. Taplin showed how to increase regular mammography use, but that may be more easily said than done with respect to racial and ethnic minorities and underserved women. Clearly the current efforts to recruit these women need improvement.

We have heard a lot about the accuracy of mammography today. It is not a perfect test. Not all cancers are found, and not all normal women have a normal test result. But it is the best test we have, and it is a pretty good test when done well.

Dramatic differences have been reported between the U.S. and other countries in the performance of screening mammography, but it is unclear whether these represent true differences in how mammograms are interpreted, or whether they relate to patient characteristics, such as age, the mix of screening and diagnostic exams, or the screening intervals, that is, is it the performance of mammography or the case mix? The data seem to show that performance has a lot to do with it.

I recently spent a year in Britain which gave me the opportunity to compare screening mammography there to that in the U.S. To examine that, I pooled data from the two countries (Smith-Bindman et al., 2003). In the case of the U.K., I used data from the National Health Service Breast Screening Program on 3.94 million women. This is a single organized screening program which differs in many ways from the United States, but basically provides mammograms very similar to those in the United States, by radiologists whose training is very similar, and with very similar technology. In the United States, I included data from two sources, 978,591 women from the Breast Cancer Surveillance Consortium and 613,388 women from the CDC's National Breast and Cervical Cancer Early Detection Program. Since performance in mammography varies by age, all the analyses I will discuss are either age-adjusted or age-stratified. Performance in mammography also varies by whether women have undergone previous mammography, so all those results are stratified by first or subsequent exams, which is made easy because in both the U.S. and the U.K. separate data are kept for the first and subsequent exams.

We looked at several measures of performance. First, what percentage of mammograms resulted in a recommendation to do a further examination, a recall

TABLE 3.3 Cancer Rates in Three Data Sets from the U.S. and the U.K.

	Mammograms	Breast Cancers	Cancers/1,000
NHSBSP	3,939,329	20,699	5.2
BCSC	978,591	4,232	4.3
CDC	613,388	2,711	4.4

for either additional non-invasive workup, a diagnostic mammogram, an ultra sound, a clinical breast exam, or a recommendation for a biopsy (open orpercutaneous), cytology or histology? I looked in particular at the rate of open surgical biopsies per 100 mammograms and likewise the rates of biopsies that did or did not discover cancer per 100 mammograms.

Balanced with the recall rate is the resulting cancer detection rate, how many cancers are you finding by calling back women. You might be willing to accept a very high recall rate if you are finding a lot of cancers, including invasive cancer and ductal carcinoma in situ (DCIS). This rate was defined as the number of cancers detected per thousand mammograms. The denominator for cancers is 1,000, whereas the denominator for recall is 100, because cancers are much rarer than recalls. I counted a cancer if it was diagnosed within one year of screening. I did not look at the false negative rate, or sensitivity, because the method of ascertainment for the three programs is very different. Furthermore, cancer detection rates have been found to very closely parallel sensitivity.

Screening is more frequent in the U.S. than in the U.K. Therefore, it is important to look at recalls and cancers detected over an interval of screening rather than at just a single screening examination. You want to know basically what would happen to a woman in any of these programs if she participated in the program over some length of time, say 10 or 20 years.

Women in the U.K. will have around three exams over a 10-year period. We examined total cancer detections and recalls assuming one first exam and several subsequent exams, and then added up the total number of recalls and cancer diagnoses. We did the same thing in the U.S., except there were more exams to add up. The cancer rate (Table 3.3) is slightly different in the different programs because of the different ages of the women screened. But we found approximately the numbers of cancers you would expect, 5 cancers per 1,000 screening examinations. In terms of the recall rate, we asked what percent of women are recalled for additional evaluation after a screening exam. The numbers are similar betweenthe two U.S. data sources. Approximately 13 or 14 percent of women are recalled for additional evaluation. In the U.K. the

TABLE 3.4 Comparative Recall Rates of Women After Mammography in the U.S. and the U.K.

Age	U.K.	BCSC	CDC
50-54	7.6	14.6	12.5
55-59	7.0	13.7	12.0
60-64	6.7	12.6	11.4

number is approximately half that (shown in Table 3.4). The recall rate is lower for subsequent exams, approximately 50 percent lower, but the trend is the same. The recall rates are twice as high in the U.S. as in the U.K.

In terms of what kinds of additional tests result from the recall exams, it turns out that most of the difference is in the non-invasive workup. In the U.S., we recall about 10 to 12 percent of women in the 50 to 54 year age group for a non-invasive further evaluation (such as ultrasound), and in the U.K it is half that at 5 percent, as shown in Table 3.5. In terms of pathologic evaluation (that is, biopsy), the numbers are much closer together.

The rates of cancer detection are very similar in the two countries. In 50- to 54-year-old women, approximately 6 cancers per 1,000 were detected by screening mammography. The cancer rate increases with age, from 6 to 12, but remains similar across the different programs. Thus, the same numbers of cancers are found despite much higher recall rates in the U.S. If we examine the data over a 10-year period for women in the 50 year age range, about 17.5 percent would be recalled in the U.K. and in the U.S., where there is a greater frequency of screening, between 40 and 50 percent. These numbers are high, but they are exactly the same as have been reported by others using different data sets. Similar results are found for women in their sixties—substantially higher recall rates in the U.S. by two- to three-fold. The numbers of cancers detected in all three data sets, however, are similar based on estimating screening over 20 years.

TABLE 3.5 Percent Recall Exams in the U.S. and the U.K.

	U.K.	BCSC	CDC
Non-Invasive Work Up			
50-54	5.3	12.8	9.3
55-59	4.6	11.8	8.8
60-64	4.1	10.5	8.1
Pathologic Evaulation			
50-54	2.4	2.3	3.2
55-59	2.3	2.3	3.1
60-64	2.5	2.5.	3.4

So in summary, the U.S. programs are really very similar, but the United States is very different from the U.K. Recall rates are twice as high in the U.S. Negative open surgical biopsy rates are two to three times as frequent in the U.S. as well. Cancer detection rates however are similar, and there is no difference in the detection of large cancers.

I can speculate that the differences between the two countries may reflect the higher rate of litigation here which is focused on delayed breast cancer diagnosis. This may lead U.S. physicians to recall patients, even when they see a finding that has a low likelihood of cancer. Additionally, in the U.K. a much smaller number of radiologists focus on screening mammography. On average their mammographers read ten times as many mammograms as their U.S. counterparts.

British radiologists know there is limited manpower. They know they can't recommend that 10 or 20 percent of women come back for diagnostic exams. There is not the capacity to handle this number, so they consciously limit recalls to the number of diagnostic mammograms they can handle. But, in fact, this is helping their program as they are finding the same numbers and types of cancers without all of the additional evaluation. Also, centralized reading and double reading are the standard, almost 100 percent, and they use this system to limit recalls.

Lastly, and I think most importantly, the U.K. has nationally set quality standards that are intensively monitored through a QA network. They have very targeted CME programs that teach radiologists to reach the standards. There are agreed-upon targets about what is desirable, that is, benchmarking. We don't have that here. We don't have a set of standards that label a recall rate of 20 percent unacceptable. On the other hand, a recall rate of two percent is not acceptable unless you find a certain number of cancers.

So they have targets, and because they have set them and because they have a coordinated effort to reach very specific recall and cancer detection rates, they are better able to reach their targets. Programs and individual physicians are subject to annual peer review. Under-performing programs and physicians are reviewed. Physicians take an exam that includes practice tests; it is voluntary, but over 90 percent of physicians take it once a year. As a result the performance of outlier radiologists has improved dramatically. There is no similar program in the U.S., but I think it might be incorporated in the programs of many health care organizations, or perhaps most easily under MQSA. I think the U.K. experience teaches us that we should focus on standardizing the interpretive components of mammography and setting performance parameters as has been done for technical performance in MQSA (summarized in Box 3.3).

BOX 3.3
Summary Recommendations

- The accuracy of screening mammography can and should be im proved. The U.K. provides an example of a success model that relies on setting clear goals and continuous quality improvement.
- Access to regular screening mammorgraphy, at least every 1-3 years should be encouraged.
- Centralizing mammography services might facilitate both goals, and would also improve the financial performance of screening.

CHARLES FINDER, M.D., Associate Director, Division of Mammography Quality and Radiation Programs, Center for Devices and Radiological Health, Food and Drug Administration: In this presentation, I will describe the historical basis for the passage of the MQSA, briefly review the current MQSA program, and outline objective indicators of program performance.

In 1985, a nationwide evaluation of X-ray trends (the NEXT study), found that there was wide variation in image quality and radiation dose among mammography facilities. In 1987, the American College of Radiology (ACR) established a voluntary program of accreditation for mammography. By July 31, 1992, 2,684 (37 percent) of the 7,246 facilities that applied had failed, which meant that only 4,662 (or about 42 percent) of the approximately 11,000 total facilities then in service were fully accredited, and there had been no on-site evaluation of these facilities. Also by 1992, only 10 states had adopted any form of legislation referable to the quality of mammography. Michigan had the most comprehensive program, which had begun in 1989. This program had equipment and personnel requirements, carried out some annual inspections, and found that 34 percent of its units failed a quality test.

These findings supported enactment of MQSA, which was signed into law on October 27, 1992, and stipulated that all mammography facilities were to be certified by October 1, 1994. The Food and Drug Administration (FDA) was tasked with developing and implementing regulations for MQSA. Interim regulations became effective on October 1, 1994. These regulations and accompanying procedures set quality standards, standards for accreditation and certification, and dealt with inspections. They closely conformed to those of the ACR. The biggest change was the initiation of annual on-site inspections.

The final regulations were implemented after a long process that dealt with notice and comment, and they went into effect on April 28, 1999. They expanded and clarified many of the interim regulations' requirements. For example, the interim regulations required that equipment be specifically designed for

mammography; in the final regulations, the FDA listed the specific requirements that the equipment had to meet.

Currently, the ACR is the only national accrediting body, and Iowa, Arkansas, and Texas are state accrediting bodies under MQSA. California withdrew as a state accrediting agency May 4, 2004, and since then accreditation in that state has been taken over by the ACR. The major function that an accrediting body performs is the review of clinical and phantom images from each mammography facility at least once every three years. Additional reviews are performed when there is a suspected public health risk. These are more intensive evaluations of a facility to determine whether or not there is a public health risk; finding such risk, FDA would go in and ask the facility to notify those patients who were at risk.

In addition to accreditation, there is also certification. Currently, the Food and Drug Administration is the only national certifier, and two states, Iowa and Illinois, are state certifiers. Issuance of MQSA certificates, which are required to lawfully provide services, is the major function of certification bodies. Certification also involves annual inspections of facilities to ensure compliance with regulatory standards. In those cases where a risk to human health has been determined, the certification agency will require the facility to notify all referring physicians and their patients that there is a problem which may require review of mammograms. This has happened several times over the course of the program, and occasionally involves as many as 10,000 patients at a time. The certifying agency provides enforcement through compliance activities, and if necessary, can impose sanctions and court actions, although this is rare.

The final regulations also set up quality standards. These cover personnel qualifications in three different categories: the interpreting physicians; the radiologic technologists; and the medical physicists. We also have standards for the reports that are sent to referring physicians and patients. All patients should now receive a lay summary of their results. There are also requirements for record retention, a medical outcomes audit, which I know many people are interested in, quality control testing, and standards for equipment and quality assurance. There is a requirement for an annual physics survey and for evaluation of equipment before it is used on patients. There are also requirements that a consumer complaint mechanism be established at all facilities and that all facilities have infection control procedures.

All interpreting physicians, radiologic technologists, and medical physicists that provide mammography services must meet specific initial and continuing training, education, and experience requirements. Specifically, the interpreting physician must have a valid state license, be either board certified in diagnostic radiology or have at least three months of formal training in mammography, have 60 category one continuing medical education (CME) credits in mammography at least 15 hours of which were obtained in the 3 years prior to qualifying

as an interpreting physician, and then have interpreted mammography examinations from 240 patients in the preceding 6-month period under the direct supervision of a qualifying interpreting physician. All interpreting physicians must have 15 CME credits within a 36-month period and must interpret 960 mammography examinations in a 24-month period.

Reporting standards require that all reports must contain an overall assessment of findings. There are requirements for communicating the results to the referring physicians and patients, and there are also requirements that the films and medical reports be retained for as much as 10 years. Reports to referring physicians must have one of the six assessment categories: negative; benign; probably benign; suspicious; highly suggestive of malignancy; or incomplete, needs additional imaging evaluation.

For the medical outcomes audit, we have a very general, some might say superficial, requirement, but it is amazing, how few facilities were even implementing this level of evaluation. All mammography facilities must have a system to follow-up all positive (suspicious or highly suggestive of malignancy) mammograms, and an audit physician must be assigned responsibility to ensure that the data are collected and analyzed on a regular facility and individual physician basis with correction of any problems identified.

The FDA requires many equipment quality assurance tests, daily, weekly, quarterly, or semiannually. There is also a requirement that a medical physicist perform a series of annual tests. These cover the equipment's basic requirements, including evaluation of the automatic exposure control, dose (which generates a lot of patient interest and concern, although problems are few), phantom image quality, and radiation output, among others. Finally, there are required tests for other mammographic modalities, meaning full field digital mammography, which has its own list of specific tests designed for that equipment.

Table 3.6 displays the data on numbers of facilities at the start of four recent fiscal years. In 2000 we had almost 10,000; as of October 1, 2003 we were down to a little over 9,100, but note that the average number of mammography units per facility has increased from 1.2 to 1.5, so actually the availability of mammography units appears to be increasing slightly.

TABLE 3.6 Mammography Facility Numbers Have Been Declining

Year	Number of Facilities
10/1/2000	9,933
10/1/2001	9,558
10/1/2002	9,306
10/1/2003	9,114

NOTE: The average number of mammography units per facility has increased from 1.2 to 1.5.

The FDA's annual inspection at each facility reviews personnel qualifications, the medical reports and lay summaries, and the outcomes audit, primarily to ensure that they have been done. We don't capture the data, but we do make sure that the facility is meeting our requirements. We check to make sure that the equipment is performing. We check dose and phantom images and processing and darkroom fog. We also check to see that the medical physicists have performed the tests as required and that there is a consumer compliance mechanism. When the inspector finds that the facility is not meeting all requirements, the facility is given an inspection observation. We have broken that down into three levels: level three is the most minor deviation, generally satisfactory; level two is facility performance that is acceptable with a deviation that may affect quality; and level one is a more significant problem, deviations that may seriously compromise quality.

The FDA began inspections under the final regulations in July 1995, and we are currently doing about 8,500 inspections each year. When we started the inspection program in 1995, only 30 percent of facilities were violation free and 2.7 percent were level one. Our data for this year through the end of April show 68.2 percent of facilities violation free and 1.9 percent at level 1. By and large, this represents a steady improvement over the years. However, there have been some bumps in the road such as when we put in new regulatory requirements like continuing personnel requirements in the late 1900s and 2000. It is noteworthy, that when the mandatory accreditation program began, and clinical images from all facilities were being reviewed, about three quarters of facilities were passing on first attempt. Now, the current percentage is about 99 percent, so there has been an objective improvement there also

A phantom is one of the ways that we evaluate image quality, not actual clinical image quality, but as a surrogate for that. The purpose of the phantom is to simulate some of the structures that we will find in a breast. The typical phantom contains 16 objects. The more objects you can see on the image, the better the image quality. Figure 3.5 displays what has happened to dose and image quality over the past two or three decades. Historically, doses were fairly high, but they have declined significantly with only a slight increase recently because breast imagers have determined that more exposed (darker) films improve image quality. Clearly, at the time doses had declined, image quality, as measured by a phantom, improved dramatically.

Those interested in more information about our program, can find it on our website at http://www.fda.gov/cdrh/mammography, and if anybody ever has any specific facility type questions, we also have a facility hotline, 1-800-838-7715 that facilities or patients can call.

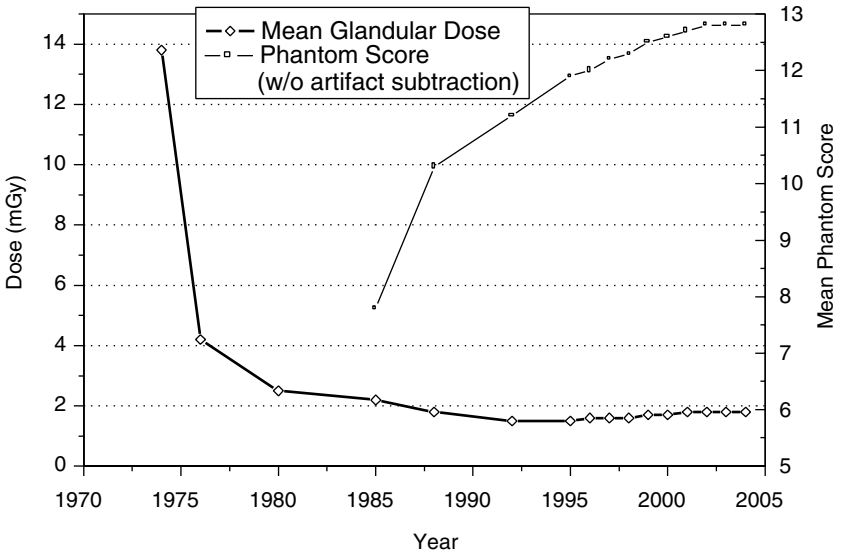

FIGURE 3.5 Image quality has improved and radiation dose decreased.

JAMES BORGSTEDE, M.D., FACR, Chairman, Board of Chancellors, American College of Radiology, Clinical Professor of Radiology, University of Colorado Health Science Center: I am a practicing radiologist in Colorado Springs who personally interprets more than 3,600 mammograms and performs more than 100 image guided breast biopsies each year. I will talk about quality and access from the perspective of the ACR and from the practitioner perspective. I commend the IOM for this report, and I stress that quality is a concern of both the IOM and the ACR; we have very few differences of opinion on how to achieve it.

Today, I will focus on four subjects: work force, liability, economics and reimbursement, and then the College's efforts. Let's talk about work force first. The U.S. General Accounting Office (GAO) recently reported on mammography capacity, and that report, which dealt with data from 1998 to 2000, gives me both some optimism for the present and some concern for the future (GAO, 2002). GAO concluded that capacity at the time of the report was adequate to deliver mammography services. Facilities had decreased during the study period by 5 percent, numbers of machines had increased by 11 percent, technologists had increased by 21 percent, and mammography had increased by 15 percent. ACR finds that these trends are continuing. In 2002-2003 the College sent a survey to 16,147 of our members, 9,048 (56 percent) responded; of those 4,924 were doing some mammography (54 percent) and 654 members (7 percent) reported being specialists in mammography.

On the other hand, population projections predict that the numbers of women at eligible ages for mammography will be growing by 1.25 million each year through 2020. Hence the reason for concern; are we going to be able to provide all of those women with mammography services in the future?

Furthermore, improving quality while simultaneously increasing access for growing numbers of women may be difficult. Others today have cited an analysis (Beam et al., 2003b) that concluded that to improve interpretation by eliminating poorly performing mammographers would likely result in an inadequate workforce. There is the potential for an inverse relationship between quality and access. As we increase quality, we have the potential to decrease access, and vice versa. I think we all need to be concerned about that for the future.

Another study that was also mentioned by others reported a 20-minute survey of radiology residents who had completed their breast imaging rotation (Bassett et al., 2003). Large majorities of the residents said they were more concerned about missing a potentially important finding on mammography than on abdominal CT, which is not an easy examination, and they endured more stress as a result. This illustrates some points that I want to talk about in a moment in the context of liability and reimbursement. People were very concerned about this issue. Stress is an important factor in interpreting mammograms.

Let's turn to non-physician prescreening, which is taken up in the IOM report. We need to ask why one would want to use prescreeners—presumably to increase accuracy and access. I believe computer assisted detection would be a more appropriate solution. Increased access should result only if there is a decrease in physician time per case so that more mammograms could be read per unit time. But I am concerned that this would lead to lesser quality.

We can examine this from another perspective, too. Use of prescreeners will decrease reimbursement because insurers will not pay the physician work rate for non-physician work. The Centers for Medicare and Medicaid Services (CMS) pays 85 percent of the M.D. rate in the Medicare fee schedule for services of physician's assistants or nurse practitioners. If a facility employed these prescreeners and held the time the same, but divided between non-physician and the physician so the latter could see more cases, total reimbursement per case would be reduced since the non-physician fraction would at a 15 percent discount. The situation would be even less economically attractive in the case of radiologic technologists whose time is worth 43 cents per minute according to the U.S. Bureau of Labor Statistics.

But non-physician prescreening raises additional questions. Will quality change? In my opinion, mammography differs from cervical cancer pap smear screening. This is not a binary procedure, where you can say it is normal or it is something else. There is a wide range of normal, and it is very difficult to distinguish normal from abnormal. Will radiologists agree to supervise in this environment? And will professional liability insurance carriers be willing to insure

with non-physician prescreeners doing part of the work? Will it save radiologist time? Will this approach potentially create a shortage of breast imaging technologists by diverting them from performing to interpreting mammograms? Will the number of diagnostic examinations change? Will these individuals, in effect, take screening cases that would have been called negative and move them into the diagnostic category resulting in radiologists working up more false positives? And most importantly, what will happen to women? What will happen to our patients? It is my opinion that prescreening will not improve access. It will take the best technologists out of the work flow, and it will not increase the productivity of radiologists.

Turning now to liability, the Physicians Insurance Association of America (PIAA) analyzed 450 paid malpractice claims involving breast cancer in 2002 (PIAA, 2002). In all of medicine, missed diagnosis of breast cancer was the number one condition for which patients filed a medical malpractice claim. Radiologists were the most frequent defendants, and this was the second most expensive condition in terms of indemnity, exceeded only by problems with deliveries and injured babies. Of those patients, 88 percent had at least one mammogram, and 80 percent of those mammograms were interpreted negative or equivocal. That does not necessarily mean that those mammograms were misinterpreted. That implies that the radiologists were at risk in those cases.

Lawsuits, even unsuccessful ones, are the reasons that radiologists are reluctant to interpret mammograms. Dr. Smith-Bindman's Figure 3.10 compared recall rates for first mammograms in the U.K. and the U.S. It would have been interesting to have an additional figure comparing numbers of attorneys. In my opinion, the threat of malpractice is one explanation for increased recalls for evaluation and more biopsies for benign disease.

Malpractice insurance premiums also appear to be higher for radiologists who interpret mammograms. Four companies quoted premiums to a Connecticut radiology practice ranging from no difference to 14, 17, and 29.5 percent lower if no mammography was done (Kaye, 2004). Similar data from Virginia were reported to me when I visited there. These premiums clearly discourage breast imaging.

I turn now to reimbursement, which is another factor that plays a role in discouraging radiologists from interpreting mammograms. CMS arrives at reimbursement by valuing the current procedural terminology, or CPT, codes that physicians use for billing in relative value units. These units are calculated according to resource costs needed to provide the services, and they are multiplied

BOX 3.4
Factors Used to Calculate Reimbursement for Physician Services in the Resource Based Relative Value Scale

- Physician work (does not include support staff)—55 percent
  - -Time
  - -Intensity
    - -Technical skill and physical effort
    - -Mental effort and judgement
    - -Stress associated with patient risk
- Practice expense—42 percent
- Professional liability insurance—3 percent

by a conversion factor to arrive at the actual dollars that are paid. Screening mammography as a service is broken down into subunits, as illustrated in Box 3.4, physician work (55 percent), practice expense (42 percent), and professional liability insurance (3 percent). Physician work which does not include support staff, is broken down further into time and intensity, and intensity is assessed by technical skill and physical effort, mental effort, and judgment, and the stress associated with patient risk.

In my opinion, there are three factors that are particularly germane to mammography and present a particular problem in valuing the service. Time is very important and, if reduced by the use of nonphysician prescreeners could result in decreased reimbursement. As for intensity and mental effort and judgment, we should recall the radiology resident survey which reported that effort and judgment in mammography exceeded that require for interpretation of abdominal CT scans. The same survey also emphasized the stress involved in this kind of work (Bassett et al., 2003).

Practice expense involves reimbursement for technologist work, the cost of equipment, and the like, and the final item is professional liability insurance. This latter, in my opinion, is more generously reimbursed for facilities than for physicians, although I believe the physicians incur more of the risk.

Table 3.7 provides an example of how Medicare reimbursement relates to cost in Colorado. The all inclusive payment for screening mammography in my practice, including physician work, practice expense, and professional liability insurance, is $83.58. My costs include $14.78 for compliance with MQSA and either $124.54 for hospital or $86.60 for office costs, all according to an ACR survey of 37 radiology practices in the spring of 2001. Clearly, this is not an attractive economic proposition.

TABLE 3.7 Costs Versus Payment in Screening Mammography in Colorado

Medicare (2004) Colorado	$83.58
*MQSA (cost of quality)	$14.78
*Hospital costs	$124.54
*Office costs	$86.60

* ACR survey, Spring 2001.
SOURCE: Medicare fee schedule and Spring 2001 ACR Mammography Cost Survey.

I would like to conclude by describing some of the College's efforts with organized radiology to improve mammography. Our efforts have been continuous for more than 25 years, working with government, the FDA, industry, and other organizations. Our efforts with the residency review committee of the American Council of Graduate Medical Education have resulted in an increase in the number of residency positions. The number of mammography units has increased, also. We will maintain quality, and we will improve access. This is a commitment from the College.

Among English speaking countries, England, Australia, Canada, New Zealand, and the U.S., there are comparable rates of mammography screening, but the 5-year survival is best in the United States, and only Australia has a slightly better mortality rate. So I think there is some cause for optimism here. My concern is for access in the future.

What should we be doing? In my opinion we need to enhance the current work force. We should use physician extenders, not for prescreening, but for hanging mammograms, contacting patients, and logistics work, and we certainly do that in our practice. Computer assisted detection is the way to go for that second pair of eyes, and we need to continue to work to improve its quality. We need to further advance the use and lower the cost of digital mammography. We also need to work on transmission of data and provide governmental incentives for manufacturers and communication system providers to enhance electronic transmission. That would simplify the transmission of full-field digital mammograms and encourage the use of centers of excellence.

There has been an increase in the number of radiology residency positions by 300 due to our efforts. We have to have relief from litigation, perhaps some sort of no-fault system as was suggested earlier. And we need appropriate reimbursement. Mammography cannot be the loss leader. It has to stand on its own. We also need to promote an environment of enthusiasm; enthusiasm by those of us performing mammography stimulates interest by residents. I believe that mammography offers tremendous research opportunities if one is interested in epidemiology, statistics, or new technologies such as MRI, telesynthesis, and the use of ultrasound.

DR. PISANO: Now is the time to open the discussion. I have some questions, but are there people from the audience who have questions as well?

DR. RICHARD WAGNER, Wisconsin Radiology Specialists: I have been practicing mammography for 25 years and have experienced unpleasant turf issues with surgeons. I think we need MQSA standards broadened to include stereotactic as well as open surgical biopsies. It is a quality issue that has to be resolved.

DR. BORGSTEDE: The College has accreditation programs in those areas. That would be something you could certainly promote.

DR. FINDER: Ultrasound biopsy is not covered by MQSA; we have no authority over ultrasound, but stereotactic biopsy certainly has been brought up, and we have been looking at that issue. I'm not sure that anything we could come up with would necessarily have affected your situation. The major factor for the FDA is if there is a problem, and is there something we can do to alleviate that problem. Many approaches to assuring quality of stereotactic biopsy have not yet been explored, such as the use of the audit; that sort of thing would probably have to be invoked in order to address the situation

DR. SMITH-BINDMAN: We do two to three times as many open surgical biopsies in the United States as they do in the U.K. Your example seems to involve the surgeons encouraging it. Perhaps we should have targets, or benchmarks, of what is desirable.

DR. BORGSTEDE: Are the surgeons doing cores?

DR. SMITH-BINDMAN: They are doing both. If they are doing exactly the same thing as radiologists are doing, that is one issue. But if they are proposing more open surgical biopsies, that could call for MQSA, or maybe another organization, to propose performance benchmarks.

DR. BORGSTEDE: We also have to be careful about adding more MQSA requirements as the way to solve turf battles. I personally would disagree with that. We are going to kill mammography programs with love if we keep adding on more and more requirements. They need to be appropriate, but they need to benefit the patient as the first priority. I would hope that I could prove that I should do the examinations because I can do them with quality. But anybody who can do them with quality should be able to do them.

DR. PISANO: Dr. Taplin, you talked a lot about how to improve the delivery system from risk assessment all the way to care after the patient was diagnosed. Similar things have been done at the University of North Carolina, and I think at other places as well, for example, the University of California, San Francisco, and Sloan-Kettering. We are salaried employees, not fee-for-service physicians, and I think that makes it easier to implement some care improvements. How do we motivate practitioners who are not in similar model systems to practice integrated health care? What are the financial incentives and disin-

centives? How do we get such systems implemented across more practices in the U.S.?

DR. TAPLIN: Certainly at my institution of the last 20 years everybody was on a salary and this affects motivations. Instead of battling for procedures, surgeons and radiologists were content to see their colleagues doing more work. So there is no doubt that the financial incentives motivate the people in a health care system.

However, I was very encouraged to hear this morning that CMS is thinking about alternative ways of reimbursement. I think those explorations and perhaps demonstration projects could be constructive. How do you fix the structure in which delivery occurs? CMS openness to beginning to think about alternative structures for reimbursement may help us.

In our demonstration, the critical part of improving quality was that the surgeons, radiologists, and others were talking together. It turns out it is pretty radical to have all those people, primary care physicians, nurses, radiologists, surgeons, all at the table at the same time, and meaningfully defining the kind of care they want to organize. I think that we need to think about more ways of reimbursing that kind of organization.

I should say also that our quality reviews and reporting occurred in the context of the quality improvement structure. It was important that we were reporting results to a committee which was responsible for the quality of care within the entire organization. That meant that the reviews, the information, and the reports were their business only and could not be discovered, including the reports of all the women who were given a negative interpretation and had a cancer within 1 year.

DR. BORGSTEDE: Speaking as a past president of a state medical board, you want to do that in a system with peer review protection, so that it is not discoverable.

DR. TAPLIN: I don't know what happens outside of our structure, whether there is also a quality improvement structure that can be set up for people in indemnity plans.

DR. D'ORSI: We know that breast cancer survival is worse in the U.K. than in the U.S.. How do you explain this in view of your data on their screening programs.

DR. SMITH-BINDMAN: Are you saying that, given the higher breast cancer mortality rates in the U.K., those data are inconsistent with our results suggesting mammography is done very well there?

DR. D'ORSI: Yes.

DR. SMITH-BINDMAN: I think looking at cancer mortality rates for a country can be quite complicated. It is difficult to compare mortality rates between the United States and the U.K. In the U.K. they usually look at survival

rates. A recent article on cancer survival rates in Europe including England, Wales, and Scotland reported that the U.K. survival rates were below those of most other European countries for most cancers whether screenable or not (Coleman et al., 2003). It is not clear why they are doing poorly, but it is a huge issue for them, and they are studying it. You would hate to pick out one cancer, for example, breast cancer, and conclude that they are doing worse than us, and therefore they cannot be doing better in mammography. I think they are doing a great job with their mammography program, and for the breast cancers they diagnose, they are doing very well in terms of finding the same proportion of small cancers as found in the U.S. So, I cannot confidently address why their survival statistics are below average.

Mortality is harder to compare. It is a more important comparison, and really has not been done. Mortality data are a little more objective. Survival data are influenced by the over diagnosis of early disease which will make the data look better even in the absence of real improvement. Thus, I just think that simple comparisons may be misleading.

DR. PISANO: I'm sorry we don't have more time to talk. We will be moving now to the wrap-up session in the other room, and there will be more time to interact over there.

Second Group Discussion
Developing and Delivering New DetectionTechnologies
Richard Bohmer, M.B.Ch.B., M.P.H., Assistant Professor, Harvard Business School and Member, Committee on Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis

MODERATOR AND RAPPORTEUR: We are going to be focusing on the development and delivery of innovative and new detection technologies. I wanted to highlight two elements of the thinking of the committee that were clearly influential in some of the report's recommendations largely because I think we are going to be talking a lot about technology assessment in its various forms this afternoon.

The first is the expectation that many of the new technologies with which we are dealing and will probably be discussing, based on this morning's conver-

sation, are more likely to be complements than substitutes. That is, they will add to the clinical armamentarium available to physicians in screening and detection of breast cancer, rather than represent wholesale replacement of any one technology.

That has some important implications. It means that with the development and introduction of more new and innovative technologies, physicians are going to have a much wider range of choices, choices that are potentially applicable to and limited to perhaps ever smaller and smaller subsets of the patient population. From a clinical practice point of view, that will substantially increase the complexity of the task ahead for clinicians. Some of those technologies will be used singly, some of them will be used in combination with other technologies in circumstances perhaps where the sequence and organization of technology use will be important

The second observation is that in some cases, technology adoption is highly context dependent. Some technologies are much easier to slip into the context of routine practice than others. There are medical technologies that tend to wreak substantial change on the organizations that are adopting them. CT scanning was perhaps one of the more famous examples of that.

We know that different organizations around the country are better or worse at undertaking the organizational and clinical process and work routine redesign that is needed in order to successfully adopt and make maximum use of a new technology. So we might expect more regional variation in the effectiveness of the use of new technologies.

Taken together, I think these two observations gave the committee a sense that various sorts of new data will need to be available to clinicians adopting new screening and diagnostic technologies. These data are likely to be both quantitative and qualitative, the latter being a class of data we are a little bit less used to using in medical practice. So there will be data not only about how well the technology performs in the absolute, which is the kind of data that goes to the FDA, but also about how the technology performs in comparison to other technologies or how technologies perform when used in concert with other technologies. The third class of data that I think will be needed is information on the appropriate organizational model or organizational design that best makes use of a new technology. These will be data on what we in the report call the deployment of technology, what kind of organizational capabilities, organizational structures, or clinical management processes will be required in order to make best use of a new screening or detection technology. The purpose of such data is to inform several decisions for clinicians and the organizations in which they work. The first is where to use a new technology and who to use it for, but also how to use a new technology and moreover, what supports need to be put in

place to be able to make the best use of that new technology, what resources, skills, processes.

So the perception of need for these data has shaped a lot of the committee's thinking. At least in my view of the report, that concern has been reflected in a number of the recommendations that we are going to be discussing this afternoon.

DANIEL SULLIVAN, M.D., Associate Director, Cancer Imaging Program, Division of Cancer Treatment and Diagnosis, NCI: My first general comment is about NIH funding policy and priority setting. The decisions at NIH are complex. To some extent the leaders of the Institutes and the Director of NIH have significant input, as do their Executive Committees with help from the staff. But those are affected by and take into consideration recommendations by external advisors and various external committees, statutory or otherwise. Since NIH is in the Executive Branch of government, there is also, of course, significant input from various individuals who allocate executive funds, and clearly, legislative authorizers and appropriaters can have a major influence.

With regard specifically to technology assessment at NIH, this morning Dr. Tunis mentioned that AHRQ has the mandate for that, but not much money. The fact that Congress created AHRQ and then followed up with relatively modest funding has sent an implicit signal to the NIH that, firstly, it is not the NIH mission to do technology assessment, and secondly, that it should not be one of NIH's high priorities. The Institutes have many potential priorities for their funds. Technology assessment does not get high on the list.

So I think in terms of moving the report's recommendations forward, one of the most helpful things that could be done, in addition to making reports like this available to NIH leaders and external advisory groups, is to get a positive signal from Congress.

With that as background, I would like to briefly discuss three messages I took from the report. First, how could federal agencies help facilitate useful multidisciplinary collaborations? This is a subject of increasing interest for a variety of groups and committees. The Biomedical Engineering Consortium (BECON) held a symposium on this subject a year ago. The report is on the website http://www.becon.nih.gov/becon_symposia.htm. Key recommendations to NIH from that report are summarized in Box 3.5. These changes are viewed as incentives, or removals of barriers, to team science. There are also recommendations in that report to academic institutions and to science publications. There are various committees and activities at NIH currently addressing all the recommendations. Similar comments have come from related reports, one of which was the IOM report, *Large-Scale Biomedical Science: Exploring Strategies for Future Research* that was released June 19, 2003 (Institute of Medicine, 2003).

BOX 3.5
Changes in NIH Policy Would Remove Barriers to Team Science

- Allow more than one Principal Investigator on individual grants.
- Allow multiple performance sites to receive appropriate indirect cost recovery.
- Develop improved funding mechanisms for team science.
- Give more attention to the special review needs of team science.

In particular, there is a subcommittee on Research in Business Models (RBM) to the Committee on Science in the Office of Science and Technology Policy at the White House which is intended to harmonize the research business approaches across federal agencies. One of the issues on their agenda is the acknowledgement of co-PIs across all federal agencies. At NIH, the BECON group is taking the lead in proposing a definition of co-PIs that RBM might propose be used by all federal agencies. There is an NIH co-PI implementation committee, which I co-chair, that will have a specific plan very soon. We hope that NIH will have the capacity to appoint true co-PIs on grants in a year.

My second subject involves better coordination between NIH and other federal agencies on technology assessment. Some of this has already been going on, and, I think, even increasing a little bit. You probably know about the ongoing trial comparing full-field digital mammography to conventional film-screen mammography (DMIST). That protocol was developed a few years ago, and planning included coordination among NIH staff, industry, FDA, and CMS, so all views could be heard and incorporated into the overall design. The trial is being carried out by the American College of Radiology Imaging Network (ACRIN), which is the cooperative group that we fund at NCI to review clinical trials in digital mammography. They are doing several related breast cancer screening studies, such as MRI and ultrasound (the latter co-funded by the Avon Foundation), and one for colon cancer screening by CT. In these smaller breast cancer studies, there actually was not very much input from other agencies. So I think there is an opportunity to do better.

A good recent example is the plan for the CT colonography study. There have been several reports with conflicting results in the last year of the potential of virtual colonoscopy to equal or approach the sensitivity and specificity of optical colonoscopy. We have been aware of those studies and recent results, in particular the work of Pickhardt and colleagues (Pickhardt et al., 2003). Prior to that report, we had organized a meeting which we held on December 9, just a few days after the publication appeared. We included extramural researchers in gastroenterology, epidemiology, and biostatistics, NCI, CMS, and FDA staff.

In fact, several staff members from CMS attended, indicating their considerable interest in the potential economic impact of virtual colonoscopy.

At that meeting, a number of very specific recommendations were made by the participants, and I want to highlight one that was of particular salience to CMS. They specifically requested that the trial be structured to allow evaluation of inter-site variability. The trial team went back and incorporated all the ideas, but on that particular issue, they debated whether there should be the same number of subjects at each site, for example, 150 or 200 subjects at each of 15 sites totaling 3,000 subjects, or whether it should be powered with the same number of polyps, the same number of suspicious polyps, or the same number of cancers. A scientific argument could be made for any of those. The final decision was to power it for the same number of cancers. This trial is now designed to accrue until each site finds 12 colon cancers. Therefore, the absolute final number of subjects is not predetermined, but should be approximately 2,500.

I think this experience provided a good example for the future of how agencies can come together and pool their respective interests to design a trial. Dr. Tunis this morning mentioned that there is a new NCI/CMS agreement. I suspect that this will help to assure that this kind of activity will occur more often.

My third subject is promotion of research over the entire spectrum of technology assessment, especially investigating how technology gets used after dissemination as opposed to the current emphasis on early feasibility and efficacy trials. This is a difficult problem, because it involves not just setting the priorities and providing money but also some tough scientific issues. There is not a generally accepted paradigm for technology assessment as there is for drug trials. Discovery, development, maturation, and dissemination stages comprise one schema.

Stage four dissemination studies are very difficult to do if you want to get a truly representative sample of all the people that are using the technology, particularly for an imaging technology, where the radiologist's interpretation is an integral part of the use of that technology. There is no database or registry that could give you information on all the radiologists that interpret chest X-rays, or all the conventional radiologists who do some conventional procedure. But mammography is the exception because all radiologists who interpret mammograms have to be certified as meeting quality standards by the FDA. When I was practicing mammography, I saw lots of films coming from other sites that I thought were poorly interpreted. Although we focus much on the quality of the image acquisition and the quality of the film, the best quality film is of no value to the patient unless it is interpreted appropriately. We generally think of this in terms of a so-called linear process of perception and cognition, first seeing it and then deciding what it is.

In the early 1990s, Beam, a biostatistician at Duke University at that time, and I decided to send a large set of mammograms to a truly random sample of

all radiologists in the country for interpretation in their usual settings (Beam et al., 1996). We found that there was a wide range of sensitivity and specificity. This variability is now generally well accepted, and that needs to be understood in terms of technology assessment.

This morning, Dr. Tunis listed the factors that CMS takes into consideration, and he specifically listed sensitivity and specificity. That is not the most appropriate method for determining the value of a technology. There is a very nice section in the IOM report about that issue and a very clear statement that says the receiver operating characteristic (ROC) curve is a better method for analysis.

For the aggregate performance of a technology, it is appropriate to determine the ROC curve for the technology. The ROC method takes into account the variability due to subjective interpretations of the image. Some of the comments this morning referred to the tradeoff between sensitivity and specificity. That may be true if you are talking about a technology in aggregate, because you will move up or down the ROC curve. It need not be the case for a single radiologist, however. I think there were some comments this morning that suggested some confusion about that: that if a radiologist increased his or her sensitivity, there would inevitably be loss of specificity. An individual can improve both sensitivity and specificity by moving to a different curve, because an individual does not necessarily have to move on the same curve.

In the report there is a section noting that signal detection theory can be applied to imaging technology. This is a true statement, but it can be misleading, because the example that is given in the report talks about finding an airplane with radar. The issue in imaging is that, although it is a signal to noise problem, the noise is not random noise but a highly structured coherent noise. So the problem would be more like looking for an airplane on a background of many other airplanes, or, the other analogy of signal to noise that people often use, looking for a polar bear in a snowstorm. But in imaging it would be like looking for a polar bear on a background of a lot of other things that look like a polar bear.

So the task is very much like the child's game, Where's Waldo. I use this as an example to illustrate that I think we need to understand the interpretive process and its implications for development of computer assisted detection (CAD) a lot better. I think it is an enormously important issue for improving mammography today.

The task in a typical picture is to find Waldo. For those of you who are a couple of years away from being five years old and may have forgotten, Waldo is identified by having a red and white striped hat, a red and white striped jersey, blue trousers, round face, glasses, and brown hair. One of the things that one might wonder is, if there are some eyes looking out of a house, is that Waldo.

This would be equivalent in mammography studies to seeing something that is completely obscured, and there is no possible way to make any statement about that. You would have to open up the house in order to get more information, which in mammographic equivalence would mean using ultrasound, MRI, and so forth.

There may be an individual who has a round face, brown hair, a hat that appears to be red, a red and white jersey partially obscured and we don't see the blue trousers at all. Our brain says, that meets enough of the criteria that it is probably Waldo. But even a 5-year-old would look at that and immediately say that it can't be Waldo because it's a girl. So how does our brain come to that decision? What I am suggesting is that this is not a clearly linear process of perception and cognition. There are still a lot of other things going on, so it is not so straightforward. I don't think we understand this process very well.

Some kids can do this task much more quickly than others; some probably have an innate ability, and some can learn to do it. If you think about the analogy here and how it could inform training radiologists, you probably wouldn't do it by giving them a lecture. You would do this in a very interactive way, because it requires developing skill.

I believe we do not train radiologists very effectively to develop a skill and use appropriate immediate feedback. I also think we give too much credit to the ability of a computer to sort this out. It is not surprising to me that there are recent reports suggesting that CAD in practice is not performing the way it did in the early trials (Zheng et al., 2004).

One of the things that we are doing to get at the issue of what agencies and industry can do collaboratively is to develop a very large database, a large imaging archive. We are doing a demonstration project with the spiral CT lung screening study to develop a database of images which will be available to help develop CAD. We think that this could be a model for public-private partnerships to develop multiple such databases for this kind of work.

So to summarize my third point, I think that research on the interpretive process is essential to the notion of evaluating technology. Radiologist training and feedback needs to be much more interactive than it is now. Studies of CAD implementation are necessary, that is, how does CAD really work when it is in the hands of radiologists, as opposed to how it works in the laboratory or in the hands of experts.

DR. NORTON: Over the years in drug development, we learned how to structure clinical trials, phases 1, 2, and 3, so that their results could change practice, even though we might ask many other kinds of scientific questions, targeting, relative efficacy, about a drug. In imaging, we seem not to have defined and universally agreed on practice changing, definitive trials, the results of which would change what people do.

DR. SULLIVAN: In a drug trial, it doesn't matter who gives the drug or how competent you are, the drug does or does not have an effect, and the trial can be designed to show that end point. In imaging trials, it very much depends on the skill and the abilities of the imager. Very large trials like our DMIST trial, will attempt to get at that by providing data that will allow us to compare the ROC curves of the two different technologies. The technology that has the higher ROC curve is the one that will be credible, that we will want to use. Of course, there is always the element of the level of comfort of practicing radiologists with a technology and what level of evidence they will finally accept to persuade them that they need to make a change.

DR. NORTON: I have been in oncology drug development trials long enough to know that when different doctors get dramatically different results, it could reflect better handling of drug toxicity; they got closer to the proper drug dose level. In any therapeutic protocol, there are clear instructions on preparing the drug, calculating the dose, handling toxicity, and the like. Those things are all subject to audit, and, as we have shown over the years, repeated auditing makes doctors better because of adherence to the protocol which not only gives more reliable trial results but better therapeutic outcome. That is what I hear is lacking here, the development of criteria for getting assessed, quantifying and measuring the human element. Our audits, our criteria, detect the doctors that say they can't do it right; they cannot follow the instructions, and they are not allowed to put patients on the protocol. In the absence of such criteria, you do get the variability that can makes it very hard to influence practice, because some doctors believe that their better personal skills will prove effective even when evidence from the trial is not there.

DR. SULLIVAN: I agree that the culture of drug development has matured to deal with those issues. The culture of diagnostic development is much less mature. However, I think improved criteria and maturity have been built into the four studies that I showed you, and it will be interesting to see how that plays out. We have not previously built in the ability to examine inter-site variability, for example.

DR. BOHMER: Related to what we have been discussing, how do radiologists tend to self-select into trials? Is there a population of radiologists who wish to participate in trials that is observably different from radiologists who might be using the outcome of the trial at some future date? Do we have a mechanism for screening the entry of radiologists into the trials?

DR. SULLIVAN: Again, this is relatively new. ACRIN has only been doing trials for about 4 years, and before that there were only a couple of isolated multi-site trials, so there is relatively little experience. I don't think anybody has looked at that issue.

DR. ESSERMAN: But even in the breast imaging trials, those participating are not general radiologists, by and large, right?

DR. SULLIVAN: That's right.

DR. ESSERMAN: Only about a third of breast cancer screening is done by breast imagers, so knowing that the trialists are mostly breast imagers, you can immediately say that they are very different. They are the dedicated and highly skilled, as it is in drug trials, the top of the group.

DR. SULLIVAN: Yes, it is going to be self-selected people who are interested and motivated to learn new interventions in a well-structured way.

DR. ESSERMAN: And you cannot disseminate those results to the general population which reinforces your point that you cannot stop with the trial. It is the same thing we found in the drug trials.

DR. BOHMER: Is it feasible to even think about deliberately recruiting different types of radiologists into future trials to get an early sense of the gap between efficacy and effectiveness prior to dissemination?

DR. SULLIVAN: We might do that through the community cancer center program at NCI. It is probably a matter of developing human resources at a limited number of sites. There is a tendency now to choose sites where there are motivated people who are willing to go through a big trial.

DR. BOHMER: And involving a diversity of radiologists might require much more involvement on behalf of the PIs, so there is a resource issue there too.

DR. SULLIVAN: In medical oncology, it is official policy that testing cancer drugs requires a medical oncologist, someone who is trained and knows how to do it. That is another difference for imaging, because specialization is not necessarily required, and that can affect results. I think finding who is qualified to use these machines should be part of the process.

DR. ESSERMAN: The first thing you want to discover is whether the technology works when it is implemented. If it doesn't work in that setting, forget it. But that is not where you stop. How you then test it or implement it generally might be in the context of registration trials where the focus is very specifically on tracking implementation and dissemination of skill sets. Maybe our concept should be that if you want to use it and be paid for it you must be part of the registration trial.

DR. KIRBY VOSBURGH, Ph.D., Associate Director, Center for Integration of Medicine and Innovative Technologies, Partners Health Care: I am a physicist by training, who spent 28 years in industrial research and development. Over the years, I worked on several breast imaging approaches, CT, MRI, and ultrasound-tagged optics. Our goal was to try to replace conventional X-ray mammography. Generally this did not pan out; it is a hard business. Today, I am representing and addressing the scientists and engineers that we might

charge to go back to their labs and get us some more effective technology to detect and characterize this difficult disease.

Since we have a significant amount of money to put into research on breast cancer screening, where do we get the best return on our investment? A logical extension of the results of our report suggests that we will probably not get it by trying to develop straight-up replacements for X-ray mammography. Several presentations today have reminded us that what we have now works well. In the IOM report, *Mammography and Beyond*, the difficulty of replacing screen-film mammography was strongly stated. A replacement has to be specific, sensitive, have the right ROC curves, be stable and inexpensive, run forever, and not be breakable, even by a Ph.D. It has to be really bulletproof. Our committee consensus has affirmed that the primary emphasis should rather be on getting every practitioner up to the best current levels using both technology (such as computer-aided diagnosis or tomosynthesis) and practice changes.

Every potential mammography replacement technology we could identify faces major challenges in competing with today's best current practices. So, a question is how many of these new ideas could we develop, recognizing, based on experience with digital mammography which required comparatively modest changes in clinical practice, that it will be at least a decade before we see them in widespread use. Many of us are frustrated by encounters with the inventor who has a great idea for a better way to detect breast cancer, tries it out on a sample of 12 women, and gets "very strong" results. The inventor then cannot understand why the idea is not immediately adopted. We should, of course, support ideas at a proof-of-concept stage, but recognize this is the beginning of a much more sophisticated process. You have to move quickly to a more complete evaluation. How will this reflect the clinical course of the disease and, ultimately, patient cohort survival at a reasonable cost to society?

In breast cancer screening, it is hard to obtain gold standards of the best possible performance, so large sample sizes are needed. Since the gold standard can be used as a basis to evaluate more than one new system, it is good to test new techniques in a multi-modality context, rather than looking at each technique in isolation. Early evaluations need to have adequate statistics and be designed in such a way that allows a decision on whether or not to scale up, but the bar for starting large-scale clinical trials should be very high. Since large-scale clinical trials will require that the technology be fixed over a long period of time, the changes in our understanding of the development and treatment of breast cancer, which are likely to continue at a rapid pace over that period, may not be accommodated. Lacking a major breakthrough that gives very high sensitivity and specificity on initial tests, validating a replacement technology for mammography will be so time-consuming that it is likely that the clinical land-

scape will change significantly and that the benefits of the trial well be diminished as a result.

An obvious way to improve conventional mammography is to increase image contrast. That gets us to the potential for physical or molecular markers to augment the detectability of disease in either a screening or diagnostic context. Contrast-enhancing agents have not been used in screening because they are expensive, and they may cause allergic reactions. Of course, at some point there may be an inexpensive and safe marker developed that lights up cancers and markedly improves sensitivity and specificity, so we should keep our eyes open and encourage biologists and imagers to talk to each other, to maintain a partnership.

A comparison of the development of virtual colonoscopy to our attempts to apply technology to X-ray diagnosis may be illuminating. In colon cancer, we have a disease process which is extremely will characterized and for which there is a very well-accepted treatment. If a polyp is bigger than, for example, five millimeters, it is excised, and that leads to a better outcome. There is not a lot of variation in how gastroenterologists evaluate and screen for polyps, so the "gold standard" is quite solid. However, the current screening procedure, optical colonoscopy, is expensive, time consuming, and not particularly beloved by patients, leading to poor compliance. It is, therefore, a good target for a high-tech approach. And, when virtual colonoscopy was proposed, a very strong consensus emerged, with all involved parties trying to develop a better system. That was an example of how progress can be made through scientist-engineer-research clinician partnerships. Unfortunately, some of the positive factors which have made virtual colonoscopy such a strong contender for clinical use are not present in mammography.

Prescreening and the consequent stratification of risk may open "niche" opportunities for novel imaging approaches that may not be suitable for broad application. We heard this in some of the presentations this morning. But prescreening has an important attribute that has not been mentioned. It may also account for the potential for effective treatment. An example of this would be the recent observation that tamoxifen chemotherapy for breast cancer may be more effective for patients with certain genetic characteristics. If you know that a patient has the potential for an effective response to chemotherapy, you may want to change your screening strategy to differentiate such women, perhaps by screening them more often. To the extent this type of correlation becomes more evident, the rationale for investments in screening and the practical targets for disease detection will be moving targets. Overall, these factors imply that care is progressing from "one size fits all" stratified only by age to individually tailored management of women at risk.

The same factors that apply to screening apply to diagnostic imaging, but you do have more resources; you can use contrast agents; you can take more

time. And you have the opportunity to integrate information and display it for the physician more effectively. However, in other medical applications, it has been found that it is not always a good idea to provide a packaged solution to physicians. It may be better to provide a richer set of information directly, and let the caregiver do the integration mentally. The optimal approach may be best established, as in many other cases, by iterative studies, with the clinicians and technologists working collegially. In this connection, high quality long-term archives of images with serum and tissue samples will be of great value.

The bottom line is that physical scientists and engineers should try to work as much as possible in concert with clinical care providers and the biologists who are studying the disease. They should recognize the power of biomedical science and informatics to improve the diagnosis and treatment of breast cancer. In this way the physical aspects of disease detection will be optimally designed and deployed to save more lives.

CAROLE FLAMM, M.D., M.P.H., Technology Evaluation Center (TEC), Blue Cross and Blue Shield Association: I will be presenting the Blue Cross Blue Shield system perspective on technology assessment. I am a radiologist, and I have been doing technology assessment for the past seven years. Dr. Baugh is going to present after me, speaking on coverage and reimbursement from the health plan perspective.

You have already heard from CMS how they look at the evidence. As outlined in Figure 3.6, the TEC process is a systematic review of the body of evidence in the literature, not performance of the primary studies. There is a formal

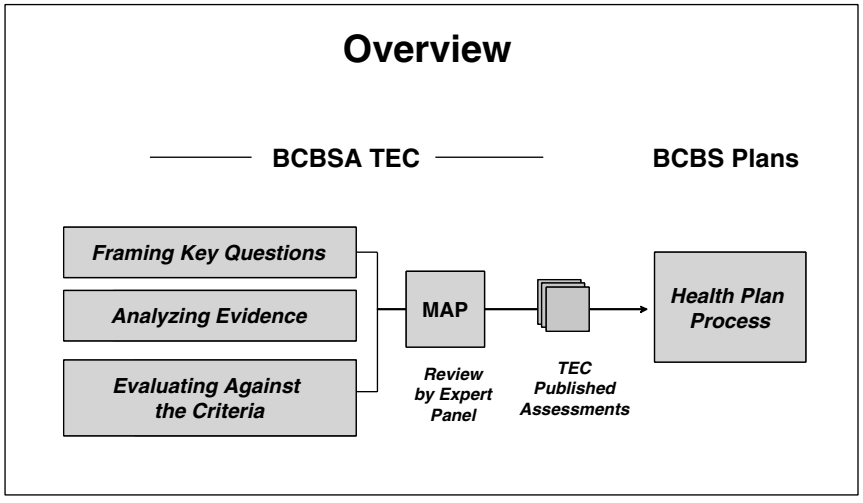

FIGURE 3.6 Technology assessment at Blue Cross and Blue Shield.

BOX 3.7
Criteria for Technology Assessment

1) The technology must have final approvale from the appropriate government regulatory bodies.
2) The scientific evidence must permit conclusions concerning the effect of the technology on health outcomes.
3) The technology must improve the net health outcome.
4) The technology must be as beneficial as any established alternatives.
5) The improvement must be attainable outside the investigational setting.

set of TEC criteria, listed in Box 3.7. I saw references to these types of criteria in the report, and CMS has similar criteria.

One of the arts of doing a technology assessment is framing the key questions, framing how you are going to look at it from a decision maker's point of view. We apply the same set of hierarchical criteria to therapeutic and diagnostic technologies, and I will discuss some of the differences in the way that plays out. The first one, FDA approval, is a necessary but not sufficient piece of information regarding effectiveness. Most technologies run into trouble on the second criterion, the sufficiency or quality of evidence. We just don't have the right kind of studies.

Then we ask if the technology improves health outcomes. Is it as beneficial as alternatives? Here, where you are talking about a technology that is going to replace another technology, it has to be as beneficial. If it is going to be used in addition to other technologies, there has to be an additional benefit, obviously.

The fifth criterion is about effectiveness versus efficacy, in other words, does the technology work in every day practice. A similar hierarchical model (illustrated in Box 3.8) of looking at the contribution of diagnostic imaging to patient management is useful in thinking about the different kinds of studies

BOX 3.8
A Hierarchical Model of Efficacy

- Level 1: Techical efficacy
- Level 2: Diagnostic accuracy efficacy
- Level 3: Diagnostic thinking efficacy
- Level 4: Therapeutic efficacy
- Level 5: Patient outcome efficacy
- Level 6: Societal efficacy

SOURCE: Fryback and Thornbury, 1991.

published (Fryback and Thornbury, 1991). Many studies look at technical quality and diagnostic accuracy, the sensitivity and specificity. I have been encouraged to see more recently studies asking how a technology changes the diagnosis, changes management, and changes outcomes. An effect on outcomes is the ultimate test we are looking for. We do not factor in cost or cost-effectiveness. The pure technology assessment is really a clinical evaluation of the evidence.

Ideally, we would like to see direct evidence, randomized controlled trials comparing outcomes with and without the test. You have heard something about the barriers to that, but it is the standard of evidence in most therapeutic technology assessments, such as drug trials. In the screening setting, we do see some randomized controlled trials and that is great, but in the diagnostic imaging literature in general, that is not the reality. Indirect evidence is the reality that links a chain of evidence: the performance of the diagnostic test; its effect on patient management; and what it does to health outcomes. What are the criteria for a positive test? Does it permit the avoidance of other tests, or invasive procedures? Does it detect a treatable condition earlier?

It is vital when using this kind of indirect framework of evidence to consider separately different patient indications. MRI of the breast differs depending of the indication, the kind of patient, the situation, and what it is being compared to. Is it a replacement for mammography, specifically for screening high risk women, or as an adjunct decision aid for biopsy in women with positive mammograms? These are different questions and require very different diagnostic performance of the test. So, the clinical context is critical.

Also, in terms of the effect on patient management, when a non-invasive test is replacing an invasive test or procedure, that represents an obvious advantage. That is an easier technology assessment question than thinking about the ultimate effect on mortality.

I'll just move on to a couple of examples. I mentioned earlier MRI of the breast in the screening setting. We have looked at this from the technology assessment point of view (http://www.bcbs.com/tec/Vol18/18_15.pdf). For women at high genetic risk, there have been studies comparing MRI and screen-film mammography (Kriege et al., 2004). In the specific population with higher than average breast density, conventional mammography is not as sensitive. Specificity is a little bit more of a tossup. But if the screening test is positive, there will be a biopsy or further workup; if the test is negative, screening will continue. The trade-off is the benefit to the true positives of earlier detection against the harms of false positives, unnecessary biopsies, and delayed diagnosis. About 6 percent of the women in a high-risk population have breast cancer. About four additional cancers will be detected for every seven additional unnecessary biopsies. That is the sensitivity tradeoff, the risk-benefit equation that is part of technology assessment.

In the high risk population, where the prevalence is high, there will be a relatively high number of false positives. Dr. Sullivan mentioned earlier that sensitivity and specificity are not the important parameters, and I agree. When comparing one test against an alternative possible replacement test, you examine the ROC curves. But particularly if you are looking for the add-on value to current management of a test, you need to understand how frequently the test is going to be called positive. What is your operating point on the ROC curve if you are going to use decision analysis modeling to figure how a positive compared to a negative test affects management, affects outcomes? So there is a little bit of that bind when we are stuck using some summary estimates of sensitivity and specificity.

The next example is positron emission tomography or PET (http://www.bcbs.com/tec/Vol18/18_14.pdf). In a woman with a positive mammogram or clinical exam who is told she needs a biopsy, can some unnecessary biopsies be avoided by using PET. A negative PET scan will spare the woman a biopsy, a positive scan will lead to biopsy. So the balance on health outcome is between the harm of delayed diagnosis versus the benefit of avoiding an unnecessary biopsy. The specificity and sensitivity are not bad, but in the populations that generated these results, there was actually a 50 percent prevalence of cancer. In such a population, there is actually a 12 percent risk of a false negative scan. So that is the way the technology assessment equation plays out in that case.

I thought these illustrations would shed some light on technology assessment in our hands. For anyone who is interested in learning more about specific technology assessments and the kind of things we do, we have a website which you can visit at http://www.bcbs.com/tec.

ERIC BAUGH, M.D., Senior Vice President, Medical Affairs, Care First Blue Cross and Blue Shield: I will discuss how we go from the technology assessment that Dr. Flamm described to coverage decisions. At Care First Blue Cross and Blue Shield we serve approximately two and a half million members in Maryland, Washington, D.C., and Delaware. We formed our own coverage policy using a variety of informational resources. A technology assessment from our Technology Evaluation Center is only one of them. The evidence for our medical policy is also reviewed by a committee of community physicians, academic experts, and plan staff. Our community is sophisticated. We have people at Johns Hopkins, the University of Maryland, George Washington, and Georgetown Hospital Center that will participate at some level in our coverage decisions.

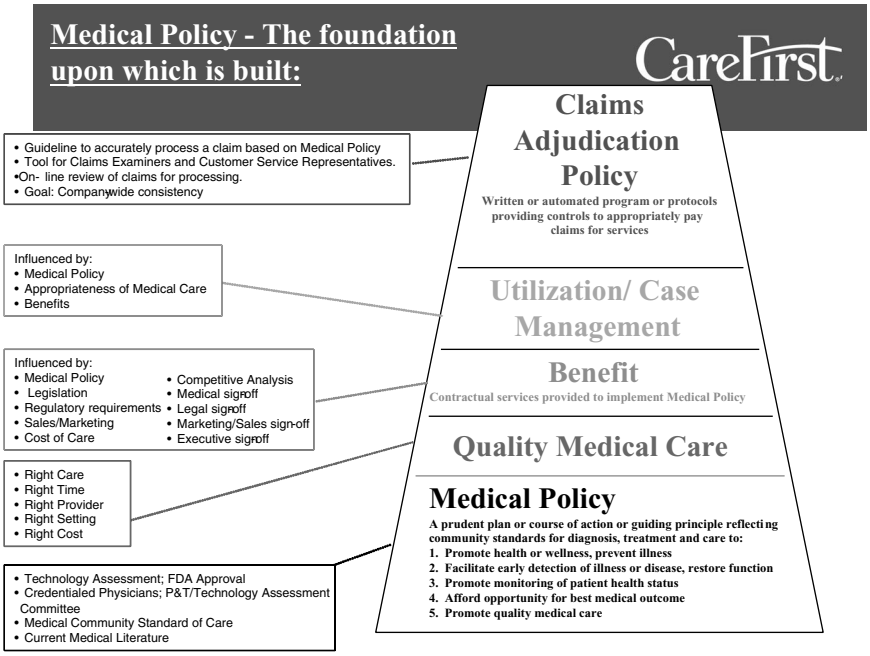

FIGURE 3.8 Care First Medical Policy.

Care First, like all Blue Cross and Blue Shield plans is an independent company and determines its own policies. Figure 3.8 illustrates our medical policy which is the foundation for everything we do. Medical policy is a proven plan or course of action or guiding principles affecting community standards of diagnosis and treatment. As you can see technology assessment, FDA approval, pharmacy and therapeutics committees, community standards of care, the medical literature, all go into helping formulate medical policy. This then determines quality of medical care, which is defined as the right care at the right time in the right setting at the right cost.

When we reach the stage of building a set of benefits, contractual services provided to implement medical policy that people can buy at a reasonable cost, cost enters into the decision on coverage. Then of course, we have utilization management. All of these filter through our claims adjudication policy as to whether or not we are going to pay for something and how much.

Medical policy development must fit contractual definitions and employ an objective standard of review and process for considering and reaching decisions.

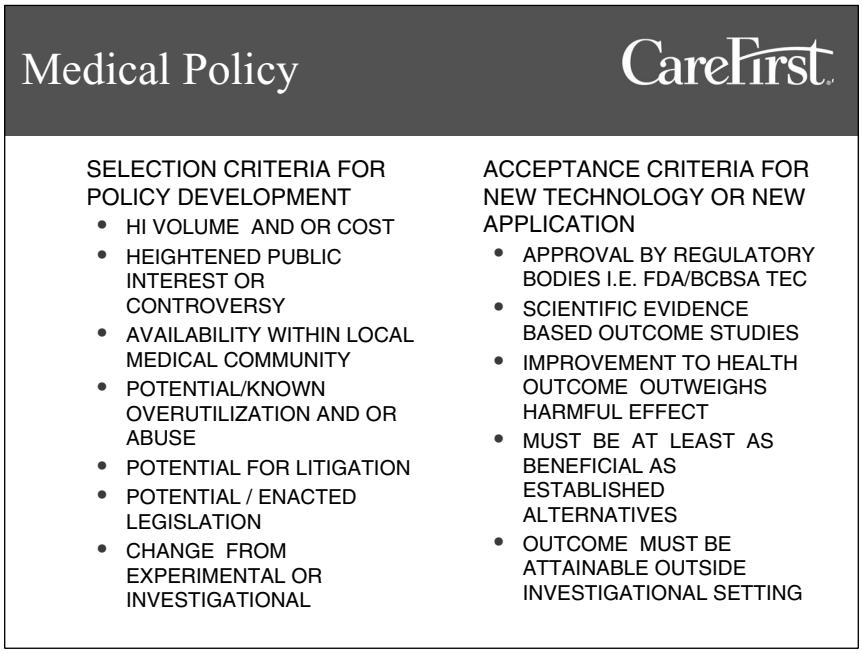

FIGURE 3.9 Coverage criteria.

We use the TEC criteria described in Figure 3.9 for determining if a new technology provides net health benefits at least as great as the best available alternative by objective evidence in peer-reviewed literature. We also use a Hayes report (http://www.hayesinc.com/) that tracks new and emerging health technologies and gives us impact utilization and cost data.

I refer to new technologies for breast cancer detection evaluated to date that provide no clinical benefit when compared to mammography or biopsy, or small benefit for a limited subset of the population when added to mammography as adjunctive. They do not substitute for existing technologies, but may add to the benefit of existing technologies for certain patients. For a new technology of this type, Care First will develop a medical or coverage policy that clearly defines for which patients and indications the technology is available. An example is MRI to investigate a woman with a positive lymph node and negative mammogram. For coverage, we must be able to verify adherence to the policy definition.

There are a number of mechanisms to implement a policy of this type. The first mechanism is prior authorization. We could require that MRI of the breast be prior authorized, and specify the documentation required before approval can be granted. This information will be reviewed by a reviewer. If the reviewer

feels that the criteria have not been met, the case will be referred to a physician reviewer. Only physicians have the authority to deny coverage.

But prior authorization programs are burdensome and unpopular. We limit the number of services that require authorizing and prefer to use other mechanisms to implement medical policies for very specific indications.

Service claims editing is a second mechanism that could be employed. Certain CPT or ICD-9 codes may be selected for review, and the claims will be separated out after service. Certain medical information will be requested from physicians to document that the indications in the policy are met. This gets back to the whole concept of evidence-based medicine and the use of protocols—how was this supposed to be used and was the doctor applying those protocols appropriately? This is burdensome to the clinician, member, and plan, and plans need to consider the time and cost of this as with other coverage restrictions.

A retrospective review after payment is a third approach to implement medical policy of this type. This gets painful if we decide that the criteria have not been met, and we ask for our money back. We look at claims experience to gauge the appropriateness of this mechanism, and apply it typically when the volume of claims is small and with limited indications. The review information will typically be used to educate participating physicians about the policy and about guidelines and protocols.

New technologies are frequently more expensive than existing technologies. PET and MRI imaging are clearly more complex than mammography. Cost is not considered in the technology assessment, but it may be a factor in formulating the coverage and payment policies, which set the coding and payment rules, the frequency limits, and payment level. Health plans need to establish payment policies when there are no existing rates, for example, for new technologies or new applications of existing technologies. The payment level may be set based on the cost of the device and the operating costs. Payors may attempt to establish a rate based on price or cost of a comparable technology, or payors may attempt to reimburse certain new technologies or drugs at the same rate as existing technologies that provide comparable clinical benefit for the condition in question. These approaches try to link price with value. They are not very popular with the manufacturers or providers of these services and could have the effect of retarding the dissemination of the technology in question.

I have tried to identify some of the selection criteria for those things that we use to establish coverage policy. These are listed in Figure 3.9. As the report for this meeting documents, FDA approval does not assure that a technology provides clinical benefit or utility. In fact, only ten percent of the new devices and tests that make it to the market have undergone trials to establish safety and effectiveness because they are cleared by the FDA through the 510-K process. It is also known that many payors will not cover a new technology that cannot dem-

onstrate an improvement in health outcomes at least as great as or better than the available alternatives.

One way to promote development of data on clinical utility after FDA approval would be for payors to provide coverage during clinical trials. Contingent on completion, such trials could establish utility for coverage. Payment only within these trials will assure their completion. In Maryland, health plans are mandated to cover patient care costs of clinical trials involving serious or life-threatening conditions. In effect, this mandate requires contingency coverage. However, these trials are not limited to those establishing clinical value.

It is important, therefore, for the health plans, other payors, and employers to interact with the research community to communicate the importance of trials that measure the impact on net health outcomes. If the trials demonstrate that the technology in question does not improve net health outcomes or is no better than conventional treatment with higher side effects, coverage will not be continued once the trial is concluded. This was our experience with autologous bone marrow transplantation for metastatic or advanced breast cancer. Coverage contingent on conduct of a trial (that is, not limited entirely to the trials) dissipates the impetus to field trials, and once provided, runs the risk of alienating members and clinicians when it must be withdrawn downstream.

DIANA PETITTI, M.D., M.P.H., Director, Research and Evaluation, Kaiser Permanente, Southern California: I am also going to speak from the decision-making payor point of view, but with the added perspective of a population and an organized system. In the United States we don't have an overall health care system, we have a non-system. I am fortunate to work within a system which defines its responsibility in terms of providing health care coverage and maximizing health insurance investment for the improvement of the health of its population.

The population served by our system comprises the 3.1 million members of the Southern California health plan—a population that is larger in size than that of New Zealand, many other countries, and 35 states.

For breast cancer, the population health perspective means that our investment of premium dollars must improve early detection for the whole population of members. Within this population health context, I am going to specifically talk about our technology assessment of CAD and our decision not to adopt and deploy this technology.

The goal for our population is to improve detection of breast cancer. We must weigh what we might spend on CAD against alternative investments of our resources. Increasing the number of women in our population who are eligible to be screened and have the test is the first competing alternative investment. Even in our system with the ability to deploy resources and outreach to our members, we have a rate of breast cancer screening in the 50- to 72-year age group of only 80 percent according to our reports in the Health Plan Employer Data and In-

formation Set (HEDIS). We consider this screening rate unacceptable given benchmarks from our other health plans of upwards of 93 percent; we believe we could attain such rates if we deployed our resources appropriately.

So in thinking through the CAD decision, it was against a backdrop of an overall screening rate of 80 percent, and a recognition that our first priority in this system would have to be to improve this rate.

However, that was not the only consideration. Within a set of possible new technologies or ways of improving performance of existing technologies, there are a number of competing approaches. Even among screened women, we have incomplete sensitivity and imperfect specificity and high false positive (suspicious and not cancer) rates, which is typical of the technology. How can we change our system and technology to improve performance of the test for the women being screened. The first way is to change the way we organize our existing services. In the IOM report, there is a case study which describes how this was accomplished in the Colorado Kaiser plan (see also Adcock, 2004). Replicating the Colorado model in our 11 facilities in Southern California is our main focus.

Now we can look carefully at CAD in the context of other alternatives. To begin our assessment, we supplemented the Blue Cross and Blue Shield evidence assessment. On reviewing this evidence, we concluded first of all that use of CAD was really not better than an experienced radiologist in terms of sensitivity. We felt that we had a pressing need to get experienced radiologists, or train a few radiologists so that they had high levels of experience, as had been achieved by the organizational changes in Colorado.

Secondly, we concluded that the evidence about the effect of CAD on callback rates in populations similar to ours and in similar broad screening efforts was poor. Evaluations of CAD had mostly been done in highly-specialized centers against specially constructed test sets. Such evaluations did not give us a very good idea of what our callback rates would be. This is important information for us because of the possible burden imposed on our system already stressed by existing service demands of about 80,000 women age-eligible for mammography and performance of about 290,000 screening mammographies each year.

And finally, at the time that we were considering CAD, we were in the process of rebuilding a number of our hospitals due to the seismic safety standards in the state of California. We were cognizant of the fact that imaging technology is moving in the direction of digital imaging, and that we would likely need to replace any CAD devices that we invested in somewhere between two and five years later. All in all, this did not seem to us a good use of resources compared to investments to increase our screening rates and better organize our services.

This kind of assessment and decision-making exemplifies why it is a privilege to work within a system. I believe it also probably represents the kind of thinking that is typical of some of the countries that have been discussed today as models. In such settings, there is a recognition that resources available to spend on any health service are fixed, and that the responsibility of decision-makers is to maximize the deployment of those resources for the benefit of the populations needing that health care service. In his presentation this morning, Dr. Tunis also suggested that we all think comprehensively in these terms about the broad clinical, economic, and social value of new or added technologies.

DR. NORTON: We heard two different views on the role of payors in doing research in imaging technology development. There is existing evidence in existing trials, or you try to get information out of ongoing trials.

DR. PETITTI: We would participate in the trial if we had that opportunity, and if it could be done for the same price as the existing service, or someone else was going to help foot the bill. I was encouraged by what Dr. Tunis said this morning. The amount of money going into direct head to head comparison, or even the trials for imaging, has been incredibly limited because the big payor is CMS, and we have a limited ability to mount them on our own.

DR. NORTON: But on the other hand, going back to autologous bone marrow transplantation for breast cancer, the Blues were paying for transplants for ten years. It took us that long to find out it did not work, and the Blues could stop paying for it. Had they supported trials, we would have gotten out in two years.

DR. BAUGH: It was not that we wanted to go in and pay for this intervention. We were mandated to pay by the courts of the United States. We got the cart before the horse.

DR. NORTON: But you wouldn't pay for participation in clinical trials.

DR. BAUGH: We would have paid for clinical trials had that been an option, but at the time, it was mandated in the courts that we pay without benefit of a clinical trial.

DR. NORTON: We could go on about that, but the point is, in terms of costs, there may be more cost-effective technology out there and we may continue to pay for technology that isn't as good.

DR. PETITTI: I think we are agreeing that someone should pay. The question is, what pocket does it come out of. So it is not so much a matter of Blue Cross, or Kaiser, or CMS paying; it is that society bears the burden of the inefficiencies that are created by using ineffective or unproven technologies or even multliple, layered technologies as seems to be happening in imaging. We need to find out how we can pay to get the evidence to sort this out.

DR. BOHMER: Or making the decision to use technologies in the absence of an evaluation of the system-wide impact of those decisions and the system-wide resources that those decisions imply. To some extent, most payors are

obliged to make one by one decisions in a way that Kaiser, because it combines payor and provider, can avoid and is better off for it.

DR. BAUGH: I agree. I think they are better off. I think the kind of evidence we are talking about needs to be gathered and paid for. People will look to find the money. We have some of the responsibility in the state of Maryland at least. When it first happened, we looked at it with mixed emotions, but I think at this point we are ready to step up to our portion.

DR. NORTON: The argument for reimbursement of the patient care costs in clinical trials is that it is cheaper in the long run for everybody, and better for patients.

DR. BAUGH: But I think it is not up to a single payor. This is something that has to be across the board and shared.

DR. PENHOET: The committee heard some strange ideas about very large trials during the course of our work. There have to be some controls if we are going to expect other people to pay for studies. And are we talking primarily about what Dr. Tunis this morning called practical clinical trials, trials focused on evidence that will help improve practice?

PARTICIPANT: And would you centralize decision-making so you have a public/private collective that could evaluate at the proof of concept point? Dr. Vosburgh, you were talking about technologies that could be out there, that could replace something that exists today, how the evidence could be collected and they could be brought to the marketplace by working with a large enough collective.

DR. PETITTI: At least from our point of view, we looked to the NIH, and maybe they become the clearinghouse, given these public-private partnerships. They have enormous credibility in the kinds of trials and the decision making process.

For example, the ALLHAT trial example from our report was an NIH trial where $20 million came from the pharmaceutical companies to pay for it. The fact that it came to us as an NIH trial with all the oversight and the integrity that implies made us willing to participate even though, I can tell you, we lost a ton of money. I have documented to the NHLBI how much money we lost in participating in that trial in the short run.

DR. BOHMER: It is an investment.

DR. PENHOET: I think the NIH review mechanisms are pretty good at sorting out the bad ideas. The issue left hanging in the air—is AHRQ a hindrance to further progress in this field or a help. It is possible that before we look at this again, we might think about refolding it back to the NIH. It is very hard given the current situation to see how it is going to work otherwise. The existing agency today, NIH, is clearly in the best position.

DR. PETITTI: I am thinking of the CT colonography trial. We would be willing to be in that trial, but it will be competitive. There will be more competent sites interested in participating than can be accommodated. That often happens with the really good NIH trials. Why would we want to be part of the trial? First of all, we get the information early. We are able to tell our members that we are looking at it. We have the satisfaction of making a public contribution, but we would probably lose money on that trial, too.

DR. VOSBURGH: Importantly, I think the NIH is taking a broader view of funding that has marketplace implications rather than scientific implications by broadening out the review process and the participation of different communities. That is essential here. It is not just a matter of clinical efficacy, but of the business case and of the potential to market it. So I think headway is being made there, but it is something that will bear attention as you move forward.

DR. PENHOET: Dr. Hanash, you never did give us your prediction of the date your proteomic marker will come to market.

DR. HANASH: Personally, I think there are multiple strategies which have merit, so which one would pan out remains to be determined. Some investment is needed for early discovery and validation to determine which markers are the winners. I think it would be very premature, at this point, to predict which particular one. In the end, I think there is going to be a continuum whereby we could start with something cheaper than a mammogram and apply progressively more expensive testing to subpopulations to confirm a diagnosis

DR. NORTON: What is the best way to collect serum proteins for proteomic analysis?

DR. HANASH: To some extent it depends on what type of marker you are after, but in terms of representing what is circulating in the blood, it is clear that plasma is best. When you subject blood to clotting, you burst a lot of cells and you activate many different subcellular systems. So what you are seeing in serum may not represent what is normally in the circulation. Plasma is a cleaner preparation because avoiding the clotting process eliminates a lot of the proteins from burst cells that you see in serum.

DR. NORTON: I am thinking about the implications of having the most informative samples. I am talking about my experiences in trials. Historically, we did not collect tissue from the tumor over the many years we were doing trials, so that, as we developed therapies that worked, we did not have samples that might identify the responsive subset of patients. Now we have the molecular technologies for classification by gene expression and gene copying and various other things that we can measure. It seems obvious to me that at some point we are also going to have protein patterns which may be informative. If we do not collect specimens prospectively during imaging or other trials, 5 years from now we will not have the opportunity to look back and identify the various subsets of

patients in which those technologies were effective. We will have missed a real golden opportunity, don't you think?

DR. HANASH: Absolutely. There is still a complete disconnect between clinical trials and molecular approaches to cancer biology. We must somehow deal with that. The cooperative trial groups do not seem very adept at designing molecular components into their studies. We have been looking for support for that without much success. At the moment, it seems that the trials are aimed only at finding out if the drug does or does not work.

DR. NORTON: We all know it is critically important. We can't get agreement on who is supposed to pay for it. Therefore, it is a question of doing something that you can afford.

DR. HANASH: We still need to figure out how to synergize molecular approaches to tumor profiling or serum profiling with cooperative trials. The NCI is very interested in having another workshop like this one to deal with that specific issue. Many challenges remain.

DR. PENHOET: I think it is worth pointing out that it is almost inevitable that in this case, as in many others, the screening test will evolve from the diagnostic test. Most of the money is going into paying for therapeutic trials and not for screening trials. Now we are finding genetic markers that predict therapies, so I think your point is well taken that money invested in clinical trial diagnostics is not money wasted in terms of eventual screening techniques.

DR. VOSBURGH: I had somewhat the same thought as Dr. Norton, but from a different perspective. This came up in early discussions of the committee and may be in our report somewhere. There is a significant role and perhaps some advocacy for the education of patients to support the acquisition of blood or tissue samples so that we can build these longitudinal databases and then go back and validate new technologies as they are developed. This is something that people can do now for the long-term advancement of detection of disease. There is a call for action here that we probably haven't emphasized as much because there are so many other good things in the report.

DR. PENHOET: It is possible that a gene chip, if you have a candidate number of genes—you would still need a few hundred—could be very inexpensive to run—or proteomics—if you only have a dozen markers or so.

DR. HANASH: We should be careful not to embellish this. That creates disappointment later. This is really a very slow painful process of an incremental nature. There is not going to be a revolution overnight; you wake up and mammography has been replaced by a 100 percent sensitive and specific test. It is incremental and very tedious.

4

Wrap-Up Session

ROGER HERDMAN, M.D., Director, National Cancer Policy Board, Institute of Medicine: I want to compliment all the speakers for a wonderful job. They have obviously prepared very carefully, and their talks were uniformly of high quality and relevant to the subject matter of this report. Our last tasks in this wrap-up session are to ask Dr. Pisano and Dr. Bohmer to review with us the bottom lines of the group discussions this afternoon, the implications for the recommendations and findings of the report, and any other relevant comments that they would like to make. Then we are very privileged to have Carolina Hinestrosa here to represent the perspective of survivors; we give the survivors the last word.

DR. PISANO: Dr. Bohmer and I are going to briefly summarize the talks in each of our group discussions because all of us at this symposium could attend only one of the concurrent sessions. Then we will try to come to some conclusions.

In our session on delivering services, we heard first from Dr. Taplin. He spoke about the organization of screening. He presented a system of improved screening services in which he had personally participated that affected both providers and patients. As an example of affecting women's behavior, he reported that women who were surveyed and discovered to have an increased risk of breast cancer because of family history were more likely to return for recommended mammograms if they were informed of their elevated risk. In the population that he studied, 52 percent of late stage breast cancers were in women who had never been screened. The multidisciplinary approach was very important to his model in which the radiologists, pathologists, surgeons, primary care physicians, and oncologists sat together and designed a system of care. Dr. Taplin's presentation raised the subject of motivating practitioners to perform within these sorts of models and how to pay for them. Dr. Tunis's comments that Medicare was interested in addressing this payment question were encouraging.

The next speaker was Dr. Smith-Bindman. She spoke about access to, and accuracy of, screening mammography and provided data comparing the U.K. to the U.S. system. She pointed out that self referrals for screening mammography probably do not reflect reality; non-whites seemed to be getting regular screening mammograms at the same rate as self referrals, but in fact, they probably are not if you look at population registries. She observed that even though the rate of recall for mammography is about 15 percent in the U.S. versus about 7 percent in Britain, it is quite variable here in different regions of the country. Actual tissue sampling is remarkably similar at about two to three percent of all women screened. In addition, in the U.S. there is a two to three time's higher rate of negative open biopsies than in the U.K.

Next we heard about the U.S. Food and Drug Administration's (FDA's) implementation of the Mammography Quality Standards Act from Dr. Finder. He presented the requirements. He mentioned that mammography facilities have indeed decreased since 2000 from 10,000 to 9,100, approximately 900 fewer facilities in a four year period. But the number of available x-ray units has actually increased slightly because of the increase in average numbers of units per facility from 1.2 to 1.5. He also described how mammography facilities have improved on MQSA inspections. Now about 70 percent of facilities do not get a violation of any type, whereas when Mammography Quality Standards Act (MQSA) first started it was about 30 percent.

Dr. Borgstede of the American College of Radiology described some practical considerations regarding improvement of mammography screening services. He referenced the 2002 U.S. General Accounting Office (GAO) study that reported that mammography capacity was adequate at that time. Because the aging of our population is increasing demand for mammography, there is concern that that report may already be out of date. In addition, he cited a study about difficulties of improving both accuracy and access. Improving accuracy by not certifying as interpreting physicians those radiologists who do not provide a good standard of accuracy will result in a reduction in manpower sufficient to impair patient access (Beam et al., 2003).

He spent quite a bit of time on our recommendation about prescreening, observing that reimbursement may suffer if we implement the recommendation. It is ironic that a second radiologist reading a mammogram is paid nothing, but a computer doing computer assisted diagonosis (CAD) generates a payment. We would hope that a trained technologist who provided a second set of eyes to read mammograms would get an additional fee. He also questioned how radiologists would supervise breast imaging technologists and posed questions that, as a practical matter, have to be answered if this recommendation is to be implemented, although in theory it is implementable right now. Current FDA regulation allows for a non-radiologist second reader. The committee envisioned extending capacity with specially trained technologists the way we do now through the use of residents and highly-qualified fellows.

Dr. Borgstede next noted that recent insurance industry data identify missed breast cancer as the number one reason for malpractice litigation in the U.S., and second only to birth-damaged infants as the most expensive condition. He concluded by providing data on mammography reimbursement, showing there is a difference between reimbursement and actual costs, especially for hospitals, less so for offices, but still negative for both.

During the discussion period, we talked about approaching Medicare for changes in reimbursement, paying for new technologies within protocols, and funding demonstration projects for some of our recommendations.

DR. BOHMER: I will report our presentations on developing and delivering new technologies out of order and begin with Dr. Vosburgh. He reminded us that mammography's sensitivity and specificity were still good. It would take quite a substantial trial spanning many years to give us sufficient evidence to justify completely replacing mammography. In the short term, therefore, new technologies are likely to be complements, not substitutes, for the existing technology. Some of the more promising advances are likely to come from the use of contrast agents or collaboration between radiologists and biologists.

We heard from Drs. Baugh and Flamm from Blue Cross and Blue Shield. They noted that they faced several kinds of insurance coverage decisions. The easy decisions were in situations in which a technology is clearly a substantial advantage for patients in all respects—then there is no difficulty in agreeing to cover it—or those technologies that add no benefit whatsoever and will not be covered. Technologies that in some respects are the same as, or better than, existing technologies are the ones that require a well defined process for medical assessment before a coverage decision can be made. Ideally, it would be nice to have a trial that compared a new technology with the incumbent technology in all respects, but often such trials do not exist. Dr. Flamm pointed out that in this situation insurers have to rely on indirect evidence, for example, (a) comparisons of a new technology's sensitivity and specificity with data on existing modalities; (b) how the performance of the new technology is likely to affect clinical decision making and, therefore; (c) the possible effect on patient outcomes. Thus, making a coverage decision in this situation can involve coordinating three sets of data.

Once the medical assessment is done, the Blues then engage in a rigorous determination of whether or not the technology should be covered, and how. At that point there are several possible strategies. The first is to cover the new technology and then apply various utilization management techniques, claims edits, or retrospective review; these need to be applied conservatively because they are burdensome and annoying to practitioners. In other circumstances, the Blues have decided to fund the clinical costs of trials to answer questions about a technology with a view to making a definitive decision at a later date.

Dr. Petitti gave us a view of a similar sort of analysis from Kaiser Permanente. This organization has the benefit of combining both the payor and the

provider function and can take a more systemic view of a technology than the Blues. She described Kaiser's recent evaluation of CAD, which resulted in a decision not to deploy CAD. The analysis examined the merits of CAD versus the benefits of alternative use of the resources for the Kaiser population.

Such benefits might include increasing breast cancer screening rates from 80 to 93 percent, reorganizing services to improve the performance of radiology (as was done in a case study from Colorado) or improving the performance of individual radiologists. Based on the available evidence, Kaiser concluded that CAD was no better than an experienced radiologist, so they would be better off training and deploying experienced radiologists than funding CAD. This context-based decision took account of system-wide concerns about the available resources within Kaiser, the kind of processes that would be more or less applicable within Kaiser, and the kinds of organizational changes that Kaiser could make. A different organization might have come to a different conclusion on reviewing the same evidence. It stands in contrast to large insurers, such as Blue Cross Blue Shield, which cannot take such local contextual factors into account in a decision about whether or not to adopt a new technology.

Our first presenter, Dr. Sullivan addressed the question of where the evidence that allowed a clearer comparison of one technology with another might come from. He cited three possibilities. First, the National Institutes of Health (NIH) is beginning to promote multidisciplinary collaboration in the conduct of trials (for example, by having more than one principal investigator on a study) so as to improve collaboration across disciplines. Second, he discussed interagency collaboration. For example, the DMIST trial involved NIH, Centers for Medicare and Medicaid Services (CMS), and FDA in the design so that the data needs of each of those three agencies would be met by this trial which compared digital with screen-film mammography. The concerns of CMS about the practical application of virtual (that is, CT) colonoscopy explain why the current National Cancer Institute (NCI) trial was designed with power to test for inter-site variability. Its aim was to address the effectiveness concerns that organizations like Kaiser worry about when deciding to adopt a new technology.

Finally, Dr. Sullivan talked about further research on new technology adoption. He described four phases in technology assessment—discovery, development, maturation, and dissemination. He pointed out that the ways to use a new screening technology are still poorly understood. Dr. Sullivan compared the problem of signal-to-noise ratios in the fields of radiology and air traffic control. In radiology, the signal is buried in other valid signals, but in air traffic control, the signal (an airplane) is buried in white noise, an entirely different problem. Hence, lessons from the airline industry may not be relevant to the problems of breast image interpretation.

At the end, the group discussed funding the kinds of trials called for in the report. We thought that funding ought to be fairly shared among payors. All payors as well as society at large realize significant benefits from effective-

ness/comparison-oriented trials. For example, an early trial of autologous bone marrow transplantation for metastatic cancer might have shown this to be a treatment not worth pursuing well in advance of the time that became understood by physicians and the public. We also observed that there is a big difference between efficacy and effectiveness, a difference not fully understood. It is an area into which many of the problems that we were convened to comment on fall. We are beginning to talk about very specific interventions to try and shrink that difference.

Final Remarks
Carolina Hinestrosa, M.A., M.P.H., Executive Vice President for Programs and Planning National Breast Cancer Coalition

The organization that I work with, the National Breast Cancer Coalition, is committed to ending breast cancer through action and advocacy. The Coalition and I think it is important to look at the report's recommendations and the problem of breast cancer from the consumer perspective.

There are many recommendations in this report, but it seems that the media have focused on the recommendations on better use of mammography. We are concerned that we are yet again reducing the approaches to breast cancer to the mammography question. As we all know, and the report says, the committee that preceded the current one, and that I was a member of, looked at mammography and concluded that it is an imperfect tool. It is the tool that we have, but it is imperfect. It is not going to solve the breast cancer problem. We are hoping that from the work of this committee, we will see more of a push to address breast cancer in a fundamental way. We believe that yes, we need to be able to find breast cancer early, but we want to know what early means and what implications that has for a patient.

It was very interesting to hear Dr. Fletcher's presentation addressing the recommendation to target women who are at high risk for breast cancer. It underscored some of the concerns I have, for example, how do we go from models of risk to telling a woman what her individual risk is? Are we going to be able to get to that point, and how useful will that be? We know that our knowledge of what causes breast cancer, what puts us at risk for breast cancer, is very limited right now. Perhaps this is one of the reasons risk stratification is so imperfect.

Furthermore, as we try to identify women and stratify them in accordance to risk, we must face a health care system that we already know is not accessible to everyone. If women are classified as high risk, will that lead to discrimination in employment or for health insurance?

Regarding, the crisis in access to mammography screening—which the previous committee also heard about—we still see that the information we have comes in bits and pieces, and what we need is a thorough and rigorous study. I urge us to be careful. I am not convinced that we have a nationwide crisis. So we welcome studying this issue, looking at all the aspects of reimbursement, of people entering the profession, of centers opening and closing, and of numbers of units, because we need to drive our policy decisions on evidence, and we need to have clear evidence of a problem.

We strongly support the committee's recommendations to look at the system of screening and move away from the ad hoc system we have, where depending on who you are, depending on whether you have insurance or not, or depending on where you live, you get screened or you don't get screened. I think that the recommendations to look at quality improvement overall, look at screening as a system, look at benchmarking, are important.

We are moving in exciting directions in breast cancer research. We are hoping, as we share the goal of saving women's lives and reducing mortality from this disease, that we put our best efforts toward going in the directions pointed out by new research findings. We have been focusing too much on a technology like mammography that has too many limitations, and we need to put our energy and resources into moving forward. We hope we can improve mammography, but let's be clear that the major impact in breast cancer is still to be made, and it is not all about mammography.

So I thank you for listening to this perspective. I want to remind you that we are expecting a lot more from you, and we will continue to push you. We hope that in this dialogue we get to our objective—to have a real impact on breast cancer.

DR. HERDMAN: I think some of our panelists and committee members are prepared to answer questions if there are any from the audience.

DR. SULLIVAN: I want to comment on benchmarking that was just mentioned in the last presentation because we need to emphasize the importance of having sensitivity data available for the population that is being looked at. I think that was not emphasized enough in the report, at least in my quick read. I think the Colorado example is useful, because I suspect they have a population registry in Colorado (Colorado is a member of the NCI Breast Cancer Screening Consortium and has a mammography registry) that gives sensitivity information, but most of what we have for benchmarking relates to specificity, that is, false positives. It is not clear from looking at changes in specificity if the radiologist's sensitivity is staying the same; actual sensitivity data are needed. Cancer yield information is a surrogate, and is helpful, but it is an imperfect surrogate.

I mention this because we know that the next Institute of Medicine report will be looking to inform the 2005 reauthorization of the MQSA, and I am sure it will come up in that discussion. It is not obligatory that an increase in specificity (fewer false positives) will come at the cost of a decrease in sensitivity for a

given radiologist, but without actually measuring and knowing sensitivity, you don't know what is happening, whether a radiologist is moving to a different ROC curve.

DR. NORTON: I think it is important to identify that reports like this are done to initiate conversation, not end conversation. I think that was the intention of this committee. I wasn't on the committee, I was an advisor to the committee, and I tried very much to stay on the periphery except when asked for advice on this. But I think it is a very hard-hitting report, and I think it is a shame that some of the public perception of it mis-identifies some of the points that were made.

Mammography clearly saves lives. I don't think we need any more studies to show that. It does not, and will not, save all lives; it is a technology that has its limitations, but a very careful review of the data shows that it has specificity and sensitivity that is very useful. If more people had access to it and it was organized better more lives would be saved and more people would take advantage of it.

There are countries that are organizing screening mammography services in an entirely different way with an impact on survival statistics and cure rates from breast cancer. We saw this morning very impressive data showing that a big difference can be made. We are all working hard on the molecular etiology and progression of breast cancer, and that is where I am focusing a lot of my energies, probably 80 percent of my research, but to throw the baby out with the bath water and say that mammography has not answered all the questions and so we should just get rid of it—I think one of the important observations in this report is that it is going to be very hard to do the trials with the resources that are, or could be, available to replace mammography.

We are all hoping that a better test will emerge and mammograms can be eliminated, but this group of experts finds that that may be extremely difficult or impossible without dramatic and unexpected changes in technology. The real advances are going to come from integrating other approaches with mammography, such as molecular diagnostics, risk assessment, and allocating patients to proper screening on that basis.

If what you have is imperfect, you have to decide whether to throw it out and replace it or to build on it. I think the conclusion is that it is something we can build on, improve its quality, improve its availability, and build on it through increasing knowledge of cancer. As far as I am concerned, it is one of the core messages of this report.

DR. HERDMAN: That was well said, Dr. Norton. We have an event like today's symposium because no report can say everything. There are different ways of looking at breast cancer research and care and different emphases that can be assigned. That is why we will report this symposium and make every effort to distribute it as widely as the original report. We consider it an important

addition or supplement that makes a difference and gets out a diverse and expanded message.

In conclusion to the day, I want to thank everybody for attending and so faithfully staying to the end. That adjourns our symposium.

References

Adcock, K. 2004. Initiative to improve mammogram interpretation. *The Permanente Journal* 8(2):12-18.

Astley, S, Mistry, T, Boggis, CRM, et al. 2002. Should we use humans or a machine to prescreen mammograms? In Peitgen, H-O, editor, *Proceedings of the Sixth International Workshop on Digital Mammography.* Heidelberg: Springer. Pp. 476-480.

Bassett, LW, Monsees, BS, Smith, RA, et al. 2003. Survey of radiology residents: breast imaging training and attitudes. *Radiology* 227(3):862-9.

Beam, CA, Layde, PM, Sullivan, DC. 1996. Variability in the interpretation of screening mammograms by U.S. radiologists. Findings from a national sample. *Arch. Intern. Med.* 156(2):209-13.

Beam, CA, Conant, EF, Sickles, EA. 2003a. Association of volume and volume-independent factors with accuracy in screening mammograms. *J. Natl Cancer Inst.* 95(4):282-90.

Beam, CA, Conant, EF, Sickles, EA, et al. 2003b. Evaluation of proscriptive health care policy implementation in screening mammography. *Radiology* 229(2):534-40.

Berlin, L, Berlin, JW. 1995. Malpractice and radiologists in Cook County, IL: trends in 20 years of litigation. *AJR Am. J. Roentgenol.* 165(4):781-8.

Bhargavan, M, Sunshine, JH. 2002. Workload of radiologists in the United States in 1998-1999 and trends since 1995-1996. *AJR Am. J. Roentgenol.* 179(5):1123-8.

Billi, JE, Wise, CG, Bills, EA, et al. 1995. Potential effects of managed care on specialty practice at a university medical center. *New Eng. J. Med.* 333(15):979-83.

Black, WC, Nease, RF, Jr., Tosteson, AN. 1995. Perceptions of breast cancer risk and screening effectiveness in women younger than 50 years of age. *J. Natl. Cancer Inst.* 87(10):720-31.

British Columbia Cancer Agency. 2003. *Screening Mammography Program of British Columbia 2001/2002 Annual Report.* Vancouver, British Columbia.

Broeders, MJM. 2004. Impact of Breast Cancer Service Screening; Facts and Uncertainties. PrintPartners, Ipskamp, Enschede, Chapter 8:105.

Burnside, E, Belkora, J, Esserman, L. 2001. The impact of alternative practices on the cost and quality of mammographic screening in the United States. *Clin. Breast Cancer* 2(2):145-52.

Coleman, MP, Gatta, G, Verdecchia, A, et al. 2003. EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. *Ann. Oncol.* 14(Suppl 5):128-49.

Cummings, SR, Duong, T, Kenyon, E, et al. 2002. Serum estradiol level and risk of breast cancer during treatment with raloxifene. *JAMA* 287(2):216-20.

Curry, SL, Taplin, SH, Anderman, C, et al. 1993. A randomized trial of the impact of risk assessment and feedback on participation in mammography screening. *Prev. Med.* 22(3):350-60.

Enzmann, DR, Anglada, P, Haviley, C, et al. 2001. Providing professional mammography services: financial analysis. *Radiology* 219(2):467-73.

Fletcher, SW, Harris, RP, Gonzalez, JJ, et al. 1993. Increasing mammography utilization: a controlled study. *J. Natl. Cancer Inst.* 85(2):112-20.

Fryback, DL, Thornbury, JM. 1991. The efficacy of diagnostic imaging. *Med. Decis. Making* 11(2):88-94.

Gail, MH, Brinton, LA, Byar, DP, et al. 1989. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J. Natl. Cancer Inst.* 81(24):1879-86.

Gilliland, FD, Rosenberg, RD, Hunt, WD, et al. 2000. Patterns of mammography use among Hispanic, American Indian, and Non-Hispanic white women in New Mexico, 1994-1997. *Am. J. Epidemiol.* 152(5):432-7.

Hillman, BJ, Fajardo, LL, Hunter, TB, et al. 1987. Mammogram interpretation by physician assistants *Am. J. Roentgenol.* 149(5):907-12.

Hunt, KA, Rosen, EL, Sickles, EA. 1999. Outcome analysis for women undergoing annual versus biennial screening mammography: a review of 24,211 examinations. *Am. J. Roentgenol.* 173(2):285-9.

Hutton, BJ, Brad, EK, Chen, JH, et al. 2004. Breast Cancer—Screening Data for Assessing Quality of Services—New York, 2000-2003. *MMWR* 53(21):455-7.

Institute of Medicine. 1990. *Breast Cancer: Setting Priorities of Effectiveness Research*. Washington, D.C.: National Academy Press.

Institute of Medicine. 1997. *A Review of the Department of Defense Program for Breast Cancer Research.* Washington, D.C.: National Academy Press.

Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, D.C.: National Academy Press.

Institute of Medicine and National Research Council. 2001. *Mammography and Beyond: Technologies for the Early Detection of Breast Cancer.* Washington, D.C.: National Academy Press.

Institute of Medicine and National Research Council. 2003. *Large-Scale Biomedical Science: Exploring Strategies for Future Research.* Washington, D.C: The National Academies Press.

Institute of Medicine and National Research Council. 2004. *Meeting Psychosocial Needs of Women with Breast Cancer.* Washington, D.C: The National Academies Press.

Institute of Medicine and National Research Council. 2005. *Saving Women's Lives, Strategies for Improving Breast Cancer Detection and Diagnosis.* Washington, D.C: The National Academies Press.

International Agency for Research on Cancer. 2002. *Breast cancer screening, Volume 7.* Lyons, France: IARC Press.

Kan, L, Olivotto, IA, Warren Burhenne, LJ, et al. 2000. Standardized abnormal interpretation and cancer detection ratios to assess reading volume and reader performance in a breast screening program. *Radiology* 215(2):563-7.

Kaye, AD. 2004. Mammography and Monkey Bars. *J. Am. Coll. Radiol.* 1(6):379-80.

Kriege, M, Brekelmans, CT, Boetes, C, et al. 2004. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *New Eng. J. Med.* 351(5):427-500.

Le Naour, F, Misek, DE, Krause, MC, et al. 2001. Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer. *Clin. Cancer Res.* 7(11):3325-7.

Meissner, HI, Smith, RA, Rimer, BK, et al. 2004. Promoting cancer screening: learning from experience. *Cancer* 101(Suppl 5):1107-17.

Michaelson, J, Satija, S, Moore, R, et al. 2002. The pattern of breast cancer screening utilization and its consequences *Cancer* 94(1):37-43.

Physician Insurers Association of America. 2002. *Breast Cancer Study: Third Edition.* Washington, D.C.

Pickhardt, PJ, Choi, JR, Hwang, I, et al. 2003. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N. Eng. J. Med.* 349(23):2191-200.

Reynolds, J. 1999. Putting prevention into practice—the community's role (news). *Ann. Int. Med.* 130(8):1707-8.

Rockhill, G, Byrne, C, Rosner, B, et al. 2003. Breast cancer risk prediction with a log-incidence model: evaluation of accuracy. *J. Clin. Epidemiol.* 56(9):856-61.

Saketkhoo, DD, Covey, AM, Sunshine, J, et al. 2002. Updated findings from a help wanted index of job advertisements and an analysis of the policy implications: is the job market shortage for diagnostic radiologist stabilizing? *Am J. Roentgenol.* 179(4):851-8.

Schwartz, D, Lellouch, J. 1967. Explanatory and pragmatic attitudes in therapeutical trials. *J. Chronic Dis.* 20(8):637-48.

Singletary, SE. 2003. Rating the risk factors for breast cancer. *Ann. Surg.* 237(4):474-82.

Smith-Bindman, R, Chu, PW, Miglioretti, D.L. et al. 2003. Comparison of screening mammography in the United States and the United Kingdom. *JAMA* 290(16): 2129-37.

Sorlie, T, Tibshirani, R, Parker, J, et al. 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. USA* 100(14):8418-23.

Stone, EG, Morton, SL, Hulscher, ME, et al. 2002. Interventions that increase use of adult immunization and cancer screening services: a meta-analysis. *Ann. Intern. Med.* 136(9):641-51.

Sumkin, JH, Klaman, HM, Graham, M, et al. 2003. Prescreening mammography by technologists: a preliminary assessment. *Am. J. Roentgenol.* 180(1):253-6.

Sunshine, JH, Cypel, YS, Schepps, G. 2002. Diagnostic radiologists in 2000: basic characteristics, practices, and issues related to the radiologist shortage. *Am. J. Roentgenol.* 178(2):291-301.

Tabar, L, Yen, MF, Bitak, B, et al. 2003. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet* 361(9367):1405-10.

Taplin, SH, Anderman, C, Grothaus, L, et al. 1994. Using physician correspondence and postcard reminders to promote mammography use. *Am. J. Public Health* 84(4): 571-4.

Taplin, SH, Barlow, WE, Ludman, E, et al. 2000. Testing reminder and motivational telephone calls to increase screening mammography: a randomized study. *J. Natl. Cancer Inst.* 92(3):233-42.

Taplin, SH, Rutler, CM, Finder, C, et al. 2002. Screening mammography: clinical image quality and the risk of interval breast cancer. *Am. J. Roentgenol.* 178(4):797-803.

Taplin, SH, Ichikawa, L, Buist, DS, et al. 2004a. Evaluating organized breast cancer screening implementation: the prevention of late stage disease? *Cancer Epidemiol. Biomarkers Prev.* 13(2):225-34.

Taplin, SH, Ichikawa, L, Ulcickas-Yood, M, et al. 2004b. A reason for late-stage breast cancer: absence of screening or detection, or breakdown in follow-up? *J. Natl. Cancer Inst*. 96(20).

Tunis, SR, Stryer, DB, Clancy, CM. 2003. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 290(12):1624-32.

U.S. General Accounting Office (GAO-02-532). 2002. *Mammography: Capacity Generally Exists to Deliver Services.* Washington, D.C.: U.S. Government Printing Office.

Wald, NJ, Hackstraw, AK, Frost, CD. 1999 When can a risk factor be used as a worthwhile screening test? *BMJ* 319(7224):1562-5.

Welch, HG. 2004. *Should I Be Tested for Cancer? Maybe Not and Here's Why.* University of California Press.

Zheng, B, Swensson, RG, Golla, S, et al. 2004. Detection and classification performance levels of mammographic masses under different computer-aided detection cueing environments. *Acad. Radiol.* 11(4):398-406.

Appendix

Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis

A Breast Cancer Research Foundation and Institute of Medicine Symposium

June 14, 2004
National Academy of Sciences
2100 C Street, N.W.
Washington, D.C.

8:30 – 9:00	Continental Breakfast in the Great Hall
9:00 – 9:15	**Plenary Session—The Lecture Room** Introduction of Symposium Evelyn H. Lauder, Founder and Chairman, The Breast Cancer Research Foundation Ed Penhoet, Ph.D., Chair, Committee on Saving Women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis and Director Science and Higher Education Programs, Gordon and Betty Moore Foundation
9:15 – 9:45	"The Pros and Cons of Screening Mammography: What Women Need to Know" An Overview of the Report's Findings on Mammography Laura Esserman, M.D., Director, Carol Franc Buck Breast Care Center and Professor of Surgery and Radiology, University of California, San Francisco

9:45 – 10:15	"Challenges to Expanding Mammography Capacity" Better Quality for Women in Screening Sites Reed Dunnick, M.D., Professor and Chair, Department of Radiology, University of Michigan
10:15 – 10:45	"Better Models for U.S. Mammography Services: Implications for Accuracy and Encouragement of Screening" Better Quality for Women Through Better Organized Mammography Robert Smith, Ph.D., Director of Cancer Screening, American Cancer Society
10:45 – 11:05	Q & A
11:05 – 11:35	"Risk Stratification for Breast Cancer Detection" Better Quality Mammography Through Better Focusing of Services Suzanne Fletcher, M.D., Professor of Ambulatory Care and Prevention, Harvard Medical School
11:35 – 12:05	"The Promise of Biomarkers in Early Detection of Breast Cancer" Better Quality for Women Through New Ways of Detecting Breast Cancer Samir Hanash, M.D., President and Chair, Human Proteome Initiative Committee and Professor, Department of Pediatrics, University of Michigan
12:05 – 12:35	"Bringing New Technologies into Service" Better Quality for Women Through New or Improved Technologies Sean Tunis, M.D., M.Sc., Chief Medical Officer, and Director, Office of Clinical Standards and Quality, Centers for Medicare and Medicaid Services
12:35 – 1:00	Q&A
1:00 – 1:45	Lunch in the Great Hall for speakers and attendees

1:45-3:30 **Simultaneous group discussions with invited speakers**

Moderators & Rapporteurs: Etta Pisano, M.D., Professor of Radiology and Biomedical Engineering, Chief of Breast Imaging and Director, Biomedical Research Imaging Center, University of North Carolina at Chapel Hill
Richard Bohmer, M.B.Ch.B., M.P.H., Assistant Professor, Harvard Business School.
Panelists: Up to 20 minute talks each, followed by group Q & A and discussion for the balance of the time.

1:45-3:30 **Delivering Better Breast Cancer Screening Services—Members Room**

To discuss the report's findings and recommendations: about enhancing mammography quality, capacity, and interpretation; concerning regulation; regarding barriers to better mammography like liability, reimbursement, and access; and about advances in mammography in every day practice.

Panelists: Stephen Taplin, M.D., Senior Scientist, Applied Research Program, NCI
Rebecca Smith-Bindman, M.D., Associate Professor, Obstetrics, Gynecology and Reproductive Medicine, Radiology and Epidemiology/Biostatistics, UCSF
Charles Finder, M.D., Associate Director, Division of Mammography Quality and Radiation Programs, FDA
James Borgstede, M.D., Chairman, Board of Chancellors, American College of Radiology

1:45- 3:30 **Developing and Delivering New Detection Technologies—Board Room**

To discuss the report's recommendations: that appropriate federal and other agencies consider ways to assemble expertise and funding toward a more comprehensive approach to the assessment, adoption, and comparison of multiple new technologies; and that sponsors should support research on technology adoption and monitoring of use in practice to identify both potential failures as well as opportunities for improvement.

Panelists: Dan Sullivan, M.D., Associate Director, Cancer Imaging Program, NCI

Kirby Vosburgh, Ph.D., Associate Director, Center for Integration of Medicine and Innovative Technologies, Partners Health Care
Carole Flamm, M.D. Technology Evaluation Center, Blue Cross Blue
Shield Association and Eric Baugh, M.D., Senior Vice President, Medical Affairs, Care First, Blue Cross Blue Shield
Diana Petitti, M.D., M.P.H., Director, Research and Evaluation, Kaiser Permanente

3:30 – 4:00 **Plenary Session—The Lecture Room**

Summary of sessions and wrap-up with rapporteurs, Etta Pisano and Richard Bohmer

4:00 Comments on the presentations and discussions from the per spective of survivors
Carolina Hinestrosa, M.A., M.P.H., Executive Vice President for Programs and Planning, National Breast Cancer Coalition

4:30 Adjourn

Acknowledgments

The Institute of Medicine acknowledges the generous support of this symposium by The Apex Foundation, The Breast Cancer Research Foundation and The Broad Reach Foundation. The Apex Foundation support was given in memory of Mabel Frost McCaw and Joan Morgan, and in honor of Sallie Nicholls, Beth Weibling, Jane Carlson Williams, Bonnie Main and Amy McGraw. Also acknowledged is the assistance of the American Cancer Society, Centers for Disease Control and Prevention, and National Cancer Institute in the planning and implementation of this symposium.

Register for the Symposium at http://www.iom.edu/event.asp?id=19062