United States General Accounting Office

GAO

Report to Congressional Committees

October 1997

MAMMOGRAPHY SERVICES

Impact of Federal Legislation on Quality, Access, and Health Outcomes

GAO/HEHS-98-11

GAO

United States
General Accounting Office
Washington, D.C. 20548

Health, Education, and
Human Services Division

B-276103

October 21, 1997

The Honorable James M. Jeffords
Chairman
The Honorable Edward M. Kennedy
Ranking Minority Member
Committee on Labor and Human Resources
United States Senate

The Honorable Thomas J. Bliley, Jr.
Chairman
The Honorable John D. Dingell
Ranking Minority Member
Committee on Commerce
House of Representatives

Breast cancer is the most commonly diagnosed nonskin cancer and the second leading cause of cancer deaths among American women. Experts estimate that during the 1990s as many as 1.8 million women will be diagnosed with breast cancer, and 500,000 will die from it. The probability of survival increases significantly, however, when the disease is discovered in its early stages. Currently, the most effective technique for early detection of breast cancer is screening mammography, an X-ray procedure that can detect small tumors and breast abnormalities up to 2 years before they can be detected by touch. The use of mammography as a tool for detecting early cancer continues to increase. The proportion of women aged 50 and older who had received mammograms in the past year increased from 26 percent in 1987 to 57 percent in 1995, according to the Centers for Disease Control and Prevention (CDC).[1]

The effectiveness of mammography as a cancer detection technique is directly related to the quality of mammography procedures. To address concerns about variations in the quality of mammography services provided by about 10,000 facilities throughout the United States and its territories, the Congress passed the Mammography Quality Standards Act of 1992 (MQSA). This act established a number of requirements aimed at strengthening mammography quality, including requiring accreditation and annual inspection of mammography facilities. The act is currently under congressional consideration for reauthorization.

[1]The proportion of women 40 to 49 who had received mammograms in the past 2 years also increased: from 59 percent in 1990 to 66 percent in 1995.

The act also mandated that we assess and report on the program established by the Food and Drug Administration (FDA) to implement these requirements. We issued two interim reports, the first focusing on the act's initial impact on access to and quality of mammography services, and the second focusing on FDA's annual inspection program.[2] As required by the act, this final report focuses on assessing the act's effect on (1) the quality of mammography services, (2) early detection of breast cancer to save lives, and (3) women's access to mammography services. We also followed up on the status of our previous recommendations to FDA (see app. II).

This report is based primarily on our analysis of data obtained from FDA's certification and annual inspection programs, as well as on research we reviewed regarding mammography's effectiveness in cancer detection. We also consulted with many mammography and cancer experts during the course of our review. Details of our scope and methodology are presented in appendix I. Our work was done between January and August 1997 in accordance with generally accepted government auditing standards.

Results in Brief

MQSA has increased mammography facilities' adherence to accepted quality assurance standards, which has, in turn, had a positive effect on mammography services. MQSA established nationwide minimum standards and required facility accreditation, which resulted in thousands of facilities' having to improve their quality assurance processes. FDA's annual inspections of facilities, now in their third year, continue to show increasing compliance with these national quality standards. Further evidence of quality improvement can be seen in the quality of the X-ray images. Before the act took effect, 11 percent of facilities tested were unable to pass image quality tests; now, the nationwide figure is 2 percent.

Experts agree that improving the quality of mammography images should lead to more accurate interpretation by physicians and, therefore, to improved early detection of breast cancer. However, neither data nor research methodologies are now in place to clearly establish these links. FDA has established federal qualification requirements for physicians who interpret mammograms but has not established criteria for measuring interpretation accuracy. Furthermore, comparable pre- and post-MQSA clinical data for measuring mammography performance and cancer outcome either do not exist or, for a number of reasons, are too limited to be useful. Some steps are being taken to address interpretation accuracy

[2]Mammography Services: Initial Impact of New Federal Law Has Been Positive (GAO/HEHS-96-17, Oct. 27, 1995) and FDA's Mammography Inspections: While Some Problems Need Attention, Facility Compliance Is Growing (GAO/HEHS-97-25, Jan. 27, 1997).

and outcome measurement issues. FDA's proposed final regulations require each facility to use its own data to monitor physicians' performance on interpretation. In addition, as provided by MQSA, the National Cancer Institute (NCI) has established a Breast Cancer Surveillance Consortium of nine research projects. These projects are making progress in both developing a methodology and collecting clinical data for assessing trends in mammography performance in detecting breast cancer and reducing mortality.

When MQSA was enacted, concern was expressed that some women might have difficulty obtaining mammography services if facilities chose to close down rather than to upgrade their operations to meet the new quality standards. We found no indication that access problems had developed as a result of MQSA. Nationwide, the number of facilities that stopped offering mammograms was nearly offset by the number of new entrants into the field. Further, 92 percent of all facilities that closed were within 5 miles of a facility that remained open, and our discussions with officials in states with the highest closure rates did not reveal any evidence that access problems had occurred.

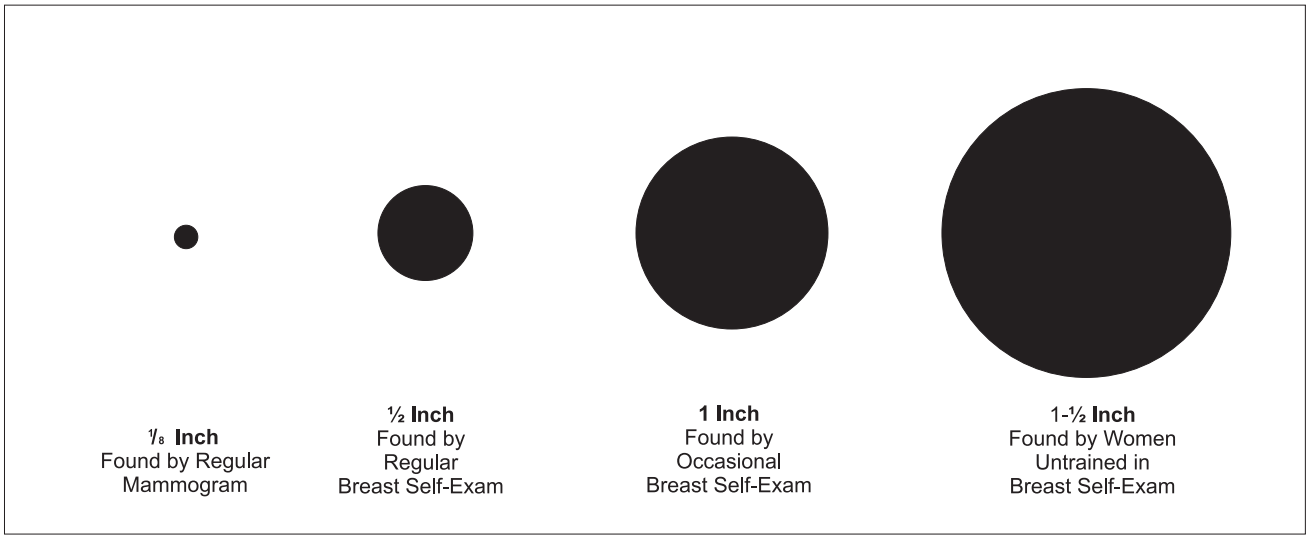
FDA has informed us that it has acted to implement our earlier recommendations for strengthening the MQSA inspection program. For example, FDA has efforts under way to establish procedures, guidance, and training to ensure timely compliance with MQSA standards. Our previous recommendations and the status of FDA actions taken on them appear in appendix II.

Background

Research studies, including randomized clinical trials, indicate that widespread use of mammography could reduce breast cancer mortality by one-third, especially in older women. The value of mammography in reducing mortality is directly tied to its ability to detect cancer at its earliest stages. Mammography is capable of detecting tumors much smaller than those detected by other means (see fig. 1).[3]

[3]Mammography, however, is not a perfect tool; even under ideal conditions, 10 to 20 percent of breast cancers cannot be detected by mammography.

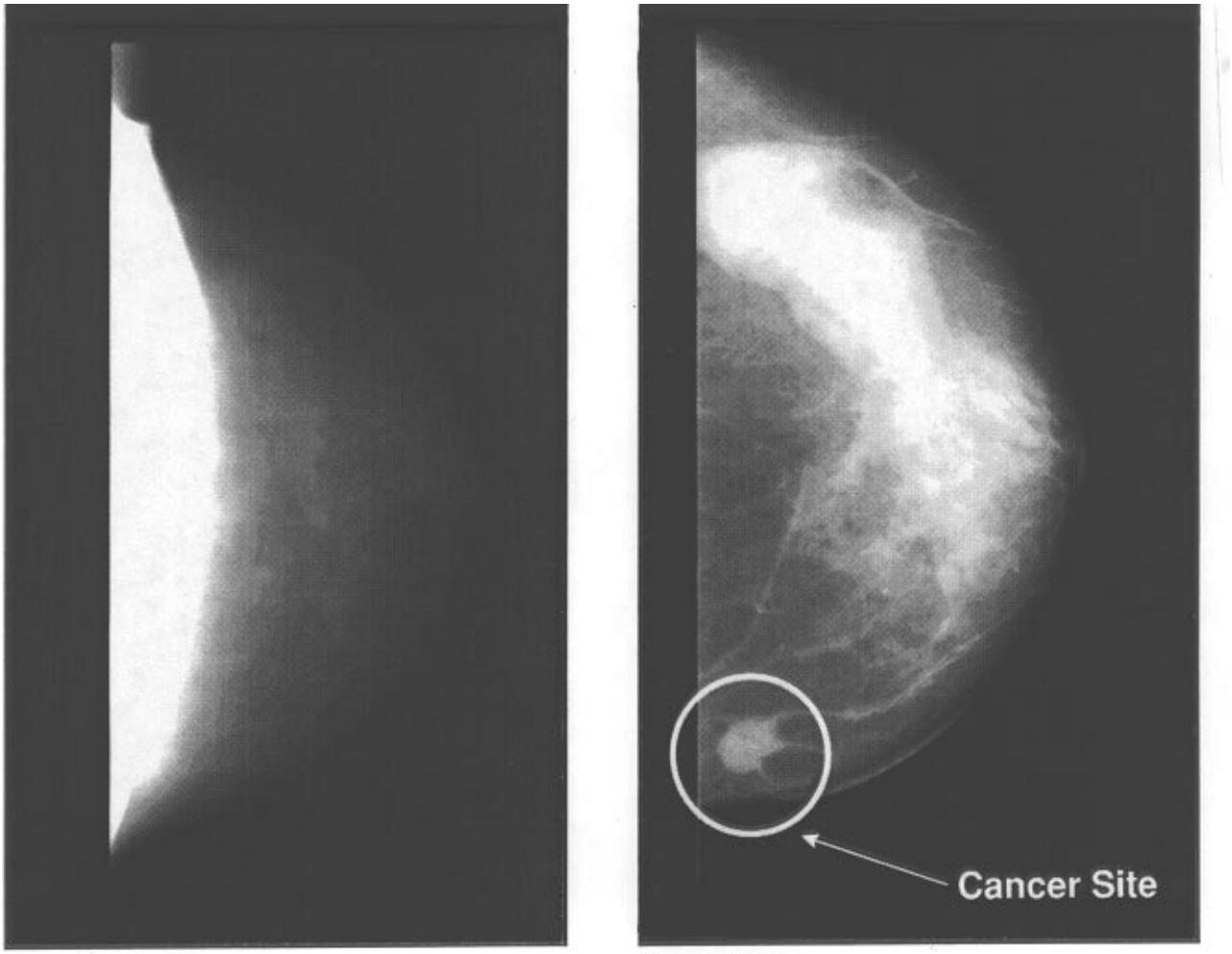
Figure 1: Size of Breast Tumors Found by Mammography and Self-Examination



Source: Altered graphic from Choices, Marion Morra and Eve Potts (New York City: Avon Books, 1994), p. 303.

Although mammography can be very useful in finding early-stage cancer, it is one of the most technically challenging radiological procedures, and ensuring the quality of the image is difficult. If the image is poor, tumors and abnormalities may go undetected. To illustrate, two images of the same patient who had a cancerous tumor are presented in figure 2. The tumor is visible in the picture on the right, where the image is of higher quality, but it is blurred and indecipherable in the picture on the left.

Figure 2: Low- and High-Quality Mammography Images of the Same Patient



Source: FDA.

Accurate interpretation of mammograms is equally as important as image quality. According to radiological experts, mammograms are the most difficult radiographic images to read, and misreading mammograms can have considerable consequences. A mammogram that is incorrectly interpreted as showing an abnormality could cause a woman to go through

unnecessary and costly follow-up procedures, such as ultrasound or biopsies. A mammogram that is read as normal when an abnormality is actually present could result in missed diagnosis of early lesions and delayed treatment, which could cost a woman's life.

MQSA established minimum national quality standards for mammography facilities and contained a number of provisions designed to ensure the quality of the image and its interpretation. Among other things, MQSA required that

- FDA establish quality standards for mammography equipment, personnel, and practices;
- all mammography facilities be accredited by an FDA-approved accrediting body once every 3 years and obtain a certificate from FDA in order to legally provide mammography services after October 1, 1994; and
- all mammography facilities have an annual evaluation by a qualified medical physicist and an annual inspection by FDA-approved inspectors that includes a test of image quality.

FDA implemented MQSA under interim regulations issued in December 1993. The quality standards of FDA's interim regulations were substantially based on standards and an accreditation program developed in 1987 by the American College of Radiology (ACR), a private, nonprofit professional association of radiologists. Since early 1994, FDA has been working with the National Mammography Quality Assurance Advisory Committee to develop final regulations.[4] On April 3, 1996, FDA published the proposed final regulations for public comment but as of September 1997 had not issued the final regulations.

MQSA Has Had a Positive Effect on the Quality of Mammography Services

To assess MQSA's effect on the quality of mammography services, we analyzed the results of implementing three important quality assurance requirements of the act: accreditation review, annual facility inspections, and testing the quality of X-ray images. Both accreditation and annual inspection processes required by MQSA show that facilities are doing a better job of complying with quality standards, and image tests show that the quality of images has improved.

[4]The committee is responsible for advising FDA on the appropriateness of quality standards for mammography facilities and accrediting bodies and for studying (1) the effect of MQSA on access to services in rural and health-professional-shortage areas, (2) the costs and benefits of compliance with MQSA, and (3) the sufficient number of medical physicists after Oct. 1, 1999, to ensure compliance with MQSA. As of September 1997, the committee had not submitted final reports to the Congress.

Accreditation Process Has Brought Improvement in Compliance With Quality Standards

MQSA required that all of the nation's approximately 10,000 mammography facilities, regardless of location or setting, pass an accreditation review to ensure that they meet quality assurance requirements for equipment performance, radiation safety, personnel qualifications, and clinical image quality, among other things. Before MQSA, between 37 and 44 percent of mammography units in the country met ACR's quality assurance requirements by participating in ACR's voluntary accreditation program. Since MQSA's implementation, FDA has approved ACR and the states of Arkansas, California, and Iowa as official accrediting bodies. Because of ACR's pre-MQSA involvement in establishing the voluntary accreditation program, it serves as the major accreditation body, responsible for more than 95 percent of the workload.

MQSA's quality standards and the related accreditation process have had a substantial effect on improving quality assurance activities. In an earlier report,[5] we noted that when MQSA initially took effect, many mammography units did not meet the standards. For example, between October 1, 1994, and August 1, 1995, about 2,600 (35 percent) of the mammography units that sought ACR accreditation initially failed to meet accreditation requirements. While almost all of these units were eventually granted full accreditation, they first had to demonstrate that they had improved their quality assurance activities.

ACR data suggest that the accreditation process continues to result in more facilities' meeting quality assurance standards. For example, the percentage of facilities that passed ACR's accreditation on the first attempt increased from 66 percent in 1995 to 82 percent in 1997.[6]

On-Site Inspections Show Continued Improvement in Compliance

MQSA's inspection requirement gives FDA another means to ensure that facilities comply with standards on a day-to-day operating level. While accreditation is generally a mail-in-process that involves the submission and review of application materials, annual inspections are conducted on

[5]GAO/HEHS-96-17, Oct. 27, 1995.

[6]The accreditation process under ACR allows facilities to go through two reviews. Facilities that fail the first accreditation review can correct deficiencies and resubmit their application materials for a second review. If a facility fails the second review, accreditation is denied. The Arkansas and Iowa accreditation bodies allow facilities to have more than two reviews before denying accreditation. While the state of California has been approved as an accreditation body, it contracts with ACR for a portion of accreditation review.

site, which allows inspectors to verify information provided during the accreditation process.[7]

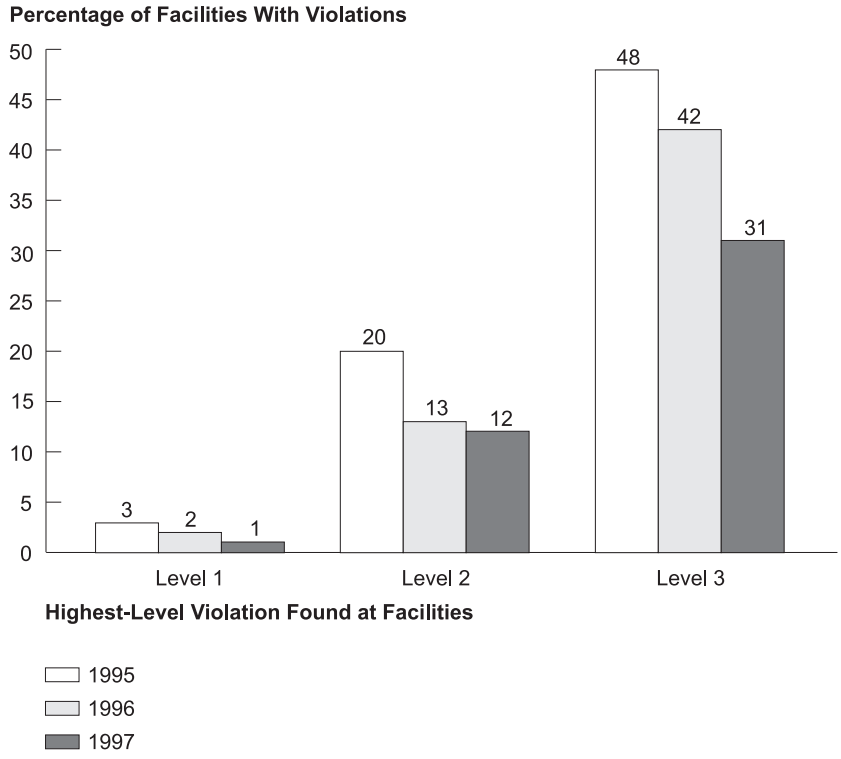
As FDA entered its third year of annual inspections, inspection results continued to show improvement in facility compliance with MQSA quality standards. Both level 1 and level 2 violations, which FDA considers to be significant and for which it requires the facility to submit written plans for corrective actions, have dropped.[8] As shown in figure 3, from fiscal year 1995 to fiscal year 1997, the proportion of inspected facilities whose highest-level violation was level 1 dropped from 3 percent to 1 percent, and the proportion whose highest-level violation was level 2 dropped from 20 percent to 12 percent.[9] Most facilities with violations had only level 3 violations (considered by FDA to be minor deviations); these numbers also dropped during the 3-year period. The portion of facilities with no violations at all nearly doubled, from about 30 percent in 1995 to about 55 percent in 1997.

[7]Annual inspections are generally performed by FDA-approved state inspectors. FDA has contracted with virtually all states and territories to conduct inspections.

[8]Level 1 violations are those that can have the most detrimental effect on the quality of mammography services—for example, a facility's having personnel who do not meet FDA's minimum qualification standards. For level 1 violations, FDA issues a warning letter; the facility must respond in writing to FDA within 15 working days of receiving the warning letter. Level 2 violations are considered less significant—for example, a facility's not having an evaluation of equipment by a medical physicist during the past 14 months—and do not generate a warning letter. However, the facility must respond in writing to FDA within 30 working days of receiving its inspection report. Level 3 violations—for example, a facility's not having the required documentation of a quality control test—do not require a written response.

[9]Under their contracts with FDA, many states began inspecting facilities in January 1995, but some did not begin until later. In fiscal year 1997, FDA began to manage its inspection workload and compare its inspection results by fiscal years. Results for 1997 are for facilities inspected during the first 9 months of the fiscal year. We have no reason to assume that the percentages will change greatly for facilities that will be inspected during the remainder of the year.

Figure 3: Comparison of Mammography Facilities' Inspection Results, Fiscal Years 1995-97



Another way to assess whether facilities are improving their compliance is to consider the extent of repeat violations—violations found in one inspection and recurring in the next inspection. As of June 30, 1997, 8,619 facilities had been inspected at least twice. Facilities had a better record for not repeating the more severe violations than they did for the minor deviations. Less than 1 percent of facilities that had level 1 violations in the first year repeated one or more of the same violations in the second year, while about 17 percent of those with level 3 violations in the first year repeated one or more of the same violations in the second year. FDA considers level 3 violations to be minor findings. However, to reduce the incidence of repeat violations, and in response to a recommendation in our second report, FDA has revised its policy guidance to include procedures for addressing repeat level 3 violations.

More Facilities Are Passing Tests of Image Quality

One of the most important aspects of FDA's annual inspection process is testing the performance of a facility's mammography system through what is called a "phantom image test." This test involves taking an X-ray image of a plastic block containing 16 test objects to determine how many of the test objects can be seen on the image generated by the facility's mammography equipment.[10] To pass the evaluation, the phantom image must show at least the minimum number of test objects required by FDA.[11]

Because FDA was already conducting some phantom image testing before MQSA, it is possible to make general comparisons about image quality before and after the law took effect. Before MQSA, FDA, in a cooperative agreement with the Conference of Radiation Control Program Directors (a professional association of state and local radiation control program officials), conducted several Nationwide Evaluation of X-ray Trends (NEXT) surveys to evaluate the technical performance of mammography facilities. The phantom image scores from these pre-MQSA NEXT surveys are considered the baseline for comparison with the phantom image scores from post-MQSA annual inspections, because the standards for evaluating phantom images are essentially identical.

NEXT survey data before MQSA implementation showed that phantom image scores were improving under ACR's voluntary accreditation program. Data from the post-MQSA inspections show continued improvement in image quality, as indicated by the acceptable rates of phantom image scores. For example, the proportion of facilities with acceptable phantom image scores has remained at 98 percent since 1995, as compared with 89 percent in 1992.

[10]The plastic block, which represents an average-sized compressed breast, contains a wax insert that holds 16 test objects—6 fibrous structures (fibers), 5 embedded microcalcifications (speck groups), and 5 different-sized tumor-like masses that simulate growths that could be cancerous.

[11]To pass the phantom image test, at least 4 fibers, 3 speck groups, and 3 masses need to be visible in the phantom image.

Data and Methodology Problems Prevent Measuring MQSA's Impact on Saving Lives

Because MQSA can be linked to improved quality of mammography images, a logical case can be made that MQSA has also helped improve mammography interpretation, thereby improving early breast cancer detection and, ultimately, saving lives. However, neither data nor research methodologies are now in place to clearly demonstrate these links. One problem is that criteria have not been established to assess how reliably images are being interpreted. Another is the lack of comparable clinical data[12] for measuring any outcome changes that have occurred as a result of the act. In addition, researchers still disagree on how to compute mammography performance and cancer outcome measures. In response to MQSA requirements, NCI has funded research and established a consortium that is making progress in collecting clinical data and resolving disagreements over how to measure mammography performance and cancer outcomes.

MQSA's Impact on the Quality of Image Interpretation Is Unknown

We were unable to determine if MQSA has had an effect on improving the accuracy and reliability with which mammograms are interpreted, because criteria are not available to determine an appropriate measure. How to develop a quality assurance system for monitoring the accuracy with which mammograms are interpreted is a controversial issue in the medical community. Several academic studies have shown wide variation in the interpretation of the same mammogram by different radiologists. On one hand, some experts suggest that some type of peer review or proficiency test is needed to improve accuracy. On the other hand, FDA and others take the position that such measures are too difficult and costly to implement and that a self-monitoring, facility-based system of physician performance assessment is a better approach to achieve the goal of quality interpretation.

FDA's current approach reflects its preference for a facility-based monitoring system that tracks and reviews mammography and pathology results. Current MQSA regulations include a general requirement that each facility have a medical audit system in place to collect and review clinical data, which includes follow-up on positive mammograms (those identified as suspicious or highly suggestive of cancer) and their correlation to biopsy results. In proposed final regulations published in April 1996, FDA requires that at least one interpreting physician review this information annually and that data be analyzed both collectively for the facility and individually for each interpreting physician. The results are then to be

[12]Clinical data include patient information such as demographics, risk factors, mammography results, and pathology results.

used to provide feedback to the interpreting physician as part of the facility's quality assurance system. However, these regulations do not include standards for measuring physician performance. FDA officials stated that one of the main reasons for the general nature of FDA's requirements is that there is no consensus in the medical and scientific community on the most desirable methodologies for arriving at such standards.

Clinical Data Are Insufficient to Assess MQSA's Contribution to Mammography Performance and Cancer Outcomes

Comparable clinical data for measuring outcome changes are not available for a number of reasons. Some data are available from clinical trials, CDC's National Breast and Cervical Cancer Early Detection Program, a facility-based medical audit system, and cancer registries. However, none of these is an appropriate source to use in assessing MQSA's effect on mammography performance and cancer outcomes.

Randomized clinical trials and other less comprehensive studies have demonstrated that regular screening mammography can significantly reduce breast cancer mortality. Such studies, while lending support to using mammography as a cancer detection technique, have not reported data that allow MQSA's specific contribution to be assessed. For example, the only clinical trial in the United States—New York's Health Insurance Plan Study—was conducted from 1963 to 1986 and was based on equipment, imaging, and interpreting techniques that are no longer applicable today.[13] Also, although several community studies have shown the effectiveness of mammography screening in early detection, the results of each study were applicable to only one geographic area, and none of the studies was current enough for a pre- and post-MQSA comparison.[14]

CDC's National Breast and Cervical Cancer Early Detection Program, established in 1991 to provide screening services to medically underserved women, has collected a fairly large volume of mammography and cancer outcome data. However, the data focus on only a subset of the population

[13]Four other major clinical trials have been undertaken in Sweden, two in the United Kingdom, and two in Canada.

[14]For example, one of the largest community studies was conducted in Albuquerque, New Mexico. The study tracked mammography and pathology results of 87,433 patients from 1991 to 1993 to assess the performance of community mammography screening. The study showed that community mammography screening could detect breast cancer at early, treatable stages and that the distribution of cancer stages was similar to that seen in successful clinical trials.

and are too incomplete to yield a meaningful analysis of the effect of MQSA on cancer outcomes.[15]

The current MQSA requirement for a facility-based medical audit system also has numerous limitations that prevent its consideration in developing national measures of MQSA's effectiveness. The fundamental limitation is that because the data collected in this process were not intended to serve as a mechanism for comparable reporting, little attention has been given to ensuring data and system consistency across facilities. Each facility decides how to collect and use the results as a quality control feedback tool, and FDA acknowledges that, as expected, there is wide variability in the data facilities include in their medical audits and in how they track results.

Problems also exist with trying to use various regional, state, and local cancer registries to evaluate MQSA's effects. These registries collect clinical data on breast cancer cases, including stage and tumor size of cancer detected. For example, the National Cancer Institute's Surveillance, Epidemiology, and End Results Program (SEER) has nine cancer registries that have been in existence since 1973.[16] Since 1992, CDC has also funded a national program that establishes or improves cancer registries in states that do not have registries such as those of the SEER program. While these registries contain clinical data that track the status of cancer cases, they do not collect data showing whether the cancer was detected by mammography. Thus, they cannot be used to measure mammography performance over time. Furthermore, while SEER cancer registry data (which contain substantial historical data) can be used to evaluate the trends of cancer incidence, cancer detection, and mortality, they are not current enough to enable a pre- and post-MQSA implementation comparison. For example, MQSA inspection of facilities began in 1995 and, at the time of our review, the SEER registry data were only current through 1994.

[15]CDC's program works with state public health agencies to provide screening, referral, and follow-up services for underserved women. The program began with 6 states in 1991 and is currently operating in 35 states. We attempted to use CDC data to compare the cancer detection rates of the program before and after MQSA, but several key data elements, such as history of prior mammograms and stage of cancers, were too incomplete to be useful for our analysis.

[16]The program consists of nine regional population-based cancer registries covering about 14 percent of the U.S. population.

No Clear Consensus Exists on Appropriate Measures for Mammography Performance and Cancer Outcomes

In addition to the problems involved with the lack of data, controversy exists over how to measure mammography performance and related cancer outcomes.

Mammography Performance Measures

Several performance measures are based on assessments of the extent to which mammography results are accurate. The two most common measures involve sensitivity and specificity.[17] Sensitivity is defined as the probability of detecting cancer when a cancer really exists, and specificity is defined as the probability of obtaining a negative mammogram when no cancer exists. The computation of these measures is shown in the following formulas:

Sensitivity = True Positive/(True Positive + False Negative)

Specificity = True Negative/(True Negative + False Positive)

Although the definitions of sensitivity and specificity are widely accepted, calculating sensitivity and specificity is complicated by the need to know the number of true and false positives, as well as the number of true and false negatives (see table 1 for definitions).

Table 1: Definitions of True and False Positive and Negative

Classification of mammogram result	Definition
True positive	Cancer diagnosed[a] during the specified follow-up period after a mammogram was interpreted as positive
False positive	No cancer diagnosed during the specified follow-up period after a mammogram was interpreted as positive
True negative	No cancer diagnosed during the specified follow-up period after a mammogram was interpreted as negative
False negative	Cancer diagnosed during the specified follow-up period after a mammogram was interpreted as negative

[a]Cancer is generally considered present only after being confirmed by a pathological diagnosis.

[17]Another type of performance measure uses the positive predictive values of mammograms (the proportion of women with a positive mammogram who are found to actually have breast cancer). This measure can be used with three separate definitions: it can be based on the number of cancers detected from positive mammograms, the number of biopsies recommended, and the number of biopsies performed. Using the second and third definitions would require information to be readily available about the number of positive mammograms that resulted in biopsy recommendations and the number of recommendations that actually resulted in biopsies' being performed.

Moreover, while researchers generally accept the above definitions, they have not agreed on the proper length of time for the follow-up period. For example, researchers have used various follow-up time periods, ranging from 7 months to over a year. Using different time periods can result in different determinations of true or false positives and true or false negatives. For example, if a mammogram is interpreted as negative and the follow-up period is 12 months, then the negative mammogram is considered a false negative if a cancer is detected during the next 12 months. However, if the specified follow-up period is 7 months, then the negative mammogram is considered a false negative only if a cancer is detected within 7 months.

In addition, researchers disagree on how certain mammography results should be classified. In 1992, the ACR developed the Breast Imaging Reporting and Data System, which established six discrete categories for interpreting physicians to use in recording the results of mammograms: "needs additional evaluation," "negative," "benign finding," "probably benign finding," "suspicious finding," and "highly suggestive of malignancy."[18] Although researchers generally agree that two categories ("negative" and "benign finding") should be considered to be negative mammograms and that two other categories ("suspicious finding" and "highly suggestive of malignancy") should be considered positive mammograms, they disagree on how the remaining categories ("needs additional evaluation" and "probably benign finding") should be classified.[19]

Standards for the follow-up time period and mammography result categories are essential for consistent application of such performance measures.

Cancer Outcome Measures

Because the ultimate goal of mammography is to reduce breast cancer mortality, some studies have used mortality reduction as a measure of mammography effectiveness in cancer outcomes. However, because women with breast cancer generally survive longer than 5 years, measuring the change in mortality reduction requires a long follow-up time. This need for a long follow-up period makes it difficult to use

[18]Although many interpreting physicians are using these categories, they are not required to do so under MQSA. However, in FDA's proposed final regulations, recording mammography results using these categories will be required for standardization purposes.

[19]The "probably benign" category is often associated with a recommendation for another mammogram within 6 months.

mortality reduction as a measure for assessing MQSA's effect on cancer outcomes, since not enough time has elapsed since MQSA's implementation.

As an alternative to mortality reduction rates, some researchers have used the early-stage disease rate as an intermediate measure of the effectiveness of mammography.[20] This intermediate measure is thought to be appropriate because early detection of breast cancer has been shown to improve survival rates. However, considerable disagreement exists about how to define early-stage cancer for measurement purposes. Perhaps the most controversial problem relates to whether or not cases of carcinoma in situ (CIS) should be included when measuring mammography effectiveness. CIS is considered the earliest stage of breast cancer and is confined to the place or site where it started. In recent years, as the number of CIS cases detected by mammography has continued to increase, researchers have questioned how to classify them. Some experts view CIS as a very slow-growing, noninvasive tumor and contend that including CIS cases as early stages of cancer will overstate the benefit of mammography in reducing mortality, since a large proportion of CIS tumors never spread. Others argue that some CIS tumors can grow quickly and develop into serious, even fatal, disease and, therefore, CIS cases should be considered early-stage cancer. Because of the above controversies, some researchers now believe that the late-stage disease rate is a more appropriate intermediate measure than the early-stage disease rate. However, there is no clear consensus on the use of this measure either.

Consortium Is Making Progress in Resolving Data and Measurement Problems

To help evaluate mammography's overall impact, MQSA called for a system that would collect a large volume of data on mammography results and cancer detection from select geographical areas. In response, NCI established the Breast Cancer Surveillance Consortium in 1994, which consists of nine federally funded research projects at sites in most of the major geographical regions of the United States.[21] These projects collect data from affiliated area mammography practices and link mammography data to cancer registry data. The Consortium's goal is to ensure that the data collected can be analyzed to address issues relating to mammography performance and cancer outcomes. By the year 2000, the Consortium

[20]The most widely used approach to categorizing cancers involves the use of staging classifications, with the size of the tumor being one of the most important factors in determining the stage of the cancer.

[21]These projects are being administered by the University of California at San Francisco, the Colorado Department of Public Health and the Environment, the Norris Cancer Center, the Fred Hutchinson Cancer Research Center, the Group Health Cooperative of Puget Sound, the University of New Mexico, the University of North Carolina, the University of Vermont, and the University of Iowa.

expects to have data on more than 3.2 million mammography examinations and more than 24,000 breast cancer cases.[22] According to Consortium members, the large sample is necessary to examine mammography performance and cancer outcomes in different health care delivery systems and regions of the country and in populations of diverse race and socioeconomic status.

To create an effective database, the Consortium is taking the following steps to overcome the kinds of data and measurement problems discussed above.

- With regard to collecting clinical data, the Consortium has spent considerable time developing a consensus on which data elements to collect in order to obtain consistent and reliable data. The Consortium has been using standardized data collection procedures and linkage mechanisms to pool data from individual projects, but 1996 was the first year common data elements were collected at all projects. The Consortium has just begun some descriptive analysis using pooled data from all projects; however, analysis of cancer outcomes will not begin until the winter of 1998, because valid data on cases of breast cancer will not be available until then.
- The Consortium is also working to obtain a consensus on performance measures such as sensitivity and specificity. Officials of the Consortium informed us that they had just begun to test aspects of these performance measures and could not tell us when they will be able to reach a consensus.

These efforts can help strengthen the quality of information about the general efficacy of mammography in improving cancer outcomes, but they are unlikely to provide substantial information about MQSA's specific effect because data are not available for a pre- and post-MQSA comparison. However, the Consortium efforts may have other positive MQSA-related benefits in that researchers involved with several of the nine research projects are currently helping mammography facilities collect data for cancer outcomes audits to meet MQSA requirements.

MQSA Did Not Limit Access

When MQSA was passed, the Congress was concerned that access to mammography services might be limited because many providers would choose to drop mammography services rather than upgrade operations to

[22]Studies show that the incidence of breast cancer ranges from 2 to 10 cancers per 1,000 mammography examinations.

comply with the standards. This has not occurred. Facility closures, both in anticipation of MQSA and since MQSA took effect, appear to have had no adverse effect on access to mammography services.

We addressed the question of the initial closures in our 1995 report. In all, 404 facilities, or about 4 percent of the approximately 10,000 facilities that were providing mammography services before MQSA was implemented, had ceased providing mammography services during the 12 months before MQSA became effective. On the basis of (1) a study conducted by a private research firm under contract with FDA and (2) our interviews with state officials, we concluded that these initial closures had no negative impact on access to services.

To assess the impact of facility closures since MQSA implementation, FDA asked the same contractor to do an updated study. The study, completed in May 1997, found that 1,085 facilities had stopped offering mammography services either temporarily or permanently between December 15, 1994, and March 19, 1997. During the same period, 922 facilities had opened or had resumed providing services. This net loss of 163 facilities since MQSA was implemented is a relatively small number when compared with the approximately 10,000 facilities operating before MQSA. Further, 99 percent of the closed facilities were located within 25 miles of another certified mammography facility, and 92 percent were within 5 miles.

To determine if the states most affected by these closures had experienced problems with access, we conducted additional follow-up work on the contractor's study. For the District of Columbia and the seven states identified in the contractor's study as having closure rates of at least 4 percent, we asked state health officials to determine if the closure of any facility on the list had caused access problems. No problems were reported to us. Officials said many of the closed facilities were either low-volume providers that did not generate enough revenue to cover the costs of meeting MQSA requirements or poor-quality providers that could not pass accreditation. Furthermore, in five of the seven states, more than 40 percent of the facilities identified as closed have actually continued to provide services in some form. For example, some facilities continued to provide services by merging with another facility or consolidating their equipment from satellite clinics. Some other facilities contracted with mobile service providers to continue serving their patients.

Conclusions

As the Congress considers reauthorization of MQSA, two points are clear with regard to assessing what the act has accomplished. First, overall, MQSA has had a positive impact on the quality of mammography services and no effect on access to them. In looking at the currently measurable areas, such as the accreditation and inspection results, the quality of X-ray images, and the extent of facility closures, the evidence is strong that the quality of services has improved and that access has not been adversely affected. We believe it is reasonable to attribute a large part of the quality improvement to (1) MQSA processes that enforced accreditation standards that were not previously followed by many facilities and (2) FDA's annual inspection process, which provides a valuable, systematic means of helping ensure that these higher standards are maintained.

Second, quantifying MQSA's effect on the accuracy of mammogram interpretation or on the improvement in cancer outcomes is much more problematic. Although data collection efforts now under way will probably make it easier to monitor the quality of image interpretation and the effects of mammography on cancer outcomes in the future, the absence of pre-MQSA data means that analysts may not be able to fully measure how the act itself has affected image interpretation and cancer outcomes.

Agency Comments

In commenting on a draft of this report, Department of Health and Human Services (HHS) officials agreed with our presentation of the issues. In addition, ACR officials provided some technical comments, which we incorporated as appropriate. Appendix III contains the full text of HHS' comments.

We are sending copies of this report to the Secretary of Health and Human Services, the Commissioner of the Food and Drug Administration, the Director of the National Cancer Institute, the Director of the Centers for Disease Control and Prevention, the Director of the Office of Management and Budget, and other interested parties. We will also make copies available to others on request.

Please contact me at (202) 512-7719 if you or your staff have any questions. Other GAO contacts and staff acknowledgments are listed in appendix IV.

Bernice Steinhardt
Director, Health Services, Quality
and Public Health Issues

Contents

Letter		1
Appendix I Scope and Methodology		24
Appendix II Follow-Up on Our January 1997 Report		26
Appendix III Comments From the Department of Health and Human Services		28
Appendix IV GAO Contacts and Staff Acknowledgments		29
Appendix V Acknowledgments		30
Tables	Table 1: Definitions of True and False Positive and Negative	14
	Table II.1: Summary of Recommendations and FDA Actions	27
Figures	Figure 1: Size of Breast Tumors Found by Mammography and Self-Examination	4
	Figure 2: Low- and High-Quality Mammography Images of the Same Patient	5
	Figure 3: Comparison of Mammography Facilities' Inspection Results, Fiscal Years 1995-97	9

Abbreviations

ACR	American College of Radiology
CDC	Centers for Disease Control and Prevention
CIS	carcinoma in situ
FDA	Food and Drug Administration
HHS	Department of Health and Human Services
MQSA	Mammography Quality Standards Act of 1992
NCI	National Cancer Institute
NEXT	Nationwide Evaluation of X-ray Trends
SEER	Surveillance, Epidemiology, and End Results Program

Scope and Methodology

To assess whether the Food and Drug Administration's (FDA) implementation of the Mammography Quality Standards Act of 1992 (MQSA) has resulted in improvement in the quality of mammography services, we relied on three analyses: the American College of Radiology's (ACR) accreditation results, FDA's annual inspection results, and FDA's pre-MQSA surveys of mammography equipment performance. Specifically, we obtained and reviewed data on ACR's accreditation failure rates and reasons for failures for each fiscal year from 1995 to 1997. In addition, we analyzed the results of FDA's annual inspections that took place between January 1995 and June 1997 at about 10,000 mammography facilities nationwide. To assess improvement in the quality of X-ray images, we compared pre- and post-MQSA phantom image scores using data from the Nationwide Evaluation of X-ray Trends (NEXT) surveys and FDA's annual inspections. The NEXT surveys collected data on the technical performance of mammography, including such things as phantom image scores, radiation dose, and existence of darkroom fog (excessive light that fogs the image). We considered the phantom image scores from pre-MQSA NEXT surveys to be the baseline for comparison with the phantom image scores from post-MQSA annual inspections. Although we did not perform a reliability assessment of FDA's data systems, we conducted several data tests and found FDA's data to be adequate to meet our objectives.

To assess whether MQSA has improved the early detection of breast cancer, we reviewed numerous studies and analyses on the accuracy of physician interpretation, breast cancer incidence and mortality, and mammography's effectiveness in cancer detection. We also interviewed many mammography and cancer experts; researchers; and officials from FDA, the National Cancer Institute (NCI), and the Centers for Disease Control and Prevention (CDC) to discuss mammography performance and cancer outcome measures, as well as the availability of current and future data to assess performance and outcome improvement. In addition, we consulted with members of NCI's Breast Cancer Surveillance Consortium who provided a written response to our questions on the methodological and data problems in measuring mammography performance and assessing MQSA's impact on cancer outcomes.

To assess MQSA's effect on accessibility, we examined an FDA contractor's study on newly opened and newly closed mammography facilities, by state, between December 1994 and March 1997. We supplemented this work by interviewing state radiation control program officials in seven states—California, Colorado, Florida, Kansas, New Hampshire, North Dakota, and Virginia—and the District of Columbia to discuss why

facilities in their localities closed and the impact that these closures had on access to mammography services.

Follow-Up on Our January 1997 Report

In our January 1997 report reviewing FDA's mammography inspection program,[23] we identified several issues that needed management attention with regard to conducting inspections and following up on deficiencies. FDA generally agreed with our recommendations and has taken action on all of them. The following table summarizes the problems, our recommendations, and the actions that FDA has taken in response.

[23]GAO/HEHS-97-25, Jan. 27, 1997.

Table II.1: Summary of Recommendations and FDA Actions

Problems we found	Our recommendations	FDA's actions
Inspection results varied considerably from state to state. It was not clear whether the differences were due to variations in the quality of mammography at different facilities or the inspectors' approaches in conducting inspections and reporting results.	Monitor the inspection results more closely to ensure consistent reporting of violations and corrections.	FDA distributed new inspection software to monitor inspection results more closely. FDA also conducted additional training of staff who audit the performance of MQSA inspectors to ensure consistent application of inspection procedures.
We found variability in how inspectors scored phantom images.	Strengthen procedures for assessing image quality by providing additional inspector training and guidance.	FDA implemented an annual program to test MQSA inspectors on phantom image scoring.
FDA had no procedure in place for clinical image review or patient notification when evidence suggested problems with the quality of a facility's mammograms.	Protect patients from risk of poor mammograms by requiring follow-up clinical image reviews and patient notification when inspections detect serious violations.	FDA established a policy that allows it to initiate a clinical image review and patient notification when serious violations are found.
When violations posed a serious health risk, FDA did not have criteria to help determine when to require the immediate suspension of a facility's certificate.	Develop procedures for (1) determining when the health risk is serious enough to justify the immediate suspension of a facility's certificate and (2) implementing the suspension.	FDA issued guidance to compliance officers to establish the criteria and procedures necessary to determine the degree of health risks and when to seek suspension of a facility's certificate.
FDA's follow-up efforts did not always ensure that corrective actions were taken on less serious violations.	Reevaluate the classification and enforcement policy for level 3 violations.	FDA revised its policy guidance to strengthen its procedures for addressing less serious violations that recur from one inspection to another.
FDA's compliance system was inadequate in three field offices included in our review because staff did not have direct access to inspection databases.	Make complete, up-to-date information on violations accessible to compliance personnel.	FDA distributed an automated compliance information system that provides current information on inspection violations.

Comments From the Department of Health and Human Services

DEPARTMENT OF HEALTH & HUMAN SERVICES

Office of Inspector General

Washington, D.C. 20201

SEP 12 1997

Ms. Bernice Steinhardt
Director, Health Services, Quality
and Public Health Issues
United States General
Accounting Office
Washington, D.C. 20548

Dear Ms. Steinhardt:

The Department has carefully reviewed your draft report entitled, "Federal Mammography Law: X-ray Quality Up, Access Unaffected, But Impact on Health Outcomes Unknown" and finds the report to be generally accurate and reflective of the Mammography Quality Standards Act program.

The Department appreciates the opportunity to comment on this draft report before its publication.

Sincerely,

June Gibbs Brown
Inspector General

The Office of Inspector General (OIG) is transmitting the Department's response to this draft report in our capacity as the Department's designated focal point and coordinator for General Accounting Office reports. The OIG has not conducted an independent assessment of these comments and therefore expresses no opinion on them.

GAO Contacts and Staff Acknowledgments

GAO Contacts

Frank Pasquier, Assistant Director, (206) 287-4861
Sophia Ku, Evaluator-in-Charge, (206) 287-4888

Staff Acknowledgments

In addition to those named above, the following individuals made important contributions to this report: Donna Bulvin, Senior Evaluator; Susan Lawes, Senior Social Science Analyst; Stan Stenersen, Senior Evaluator; Evan Stoll, Computer Specialist; and Craig Winslow, Senior Attorney.

Acknowledgments

We would like to acknowledge the members of the NCI's Breast Cancer Surveillance Consortium and officials from CDC's Division of Cancer Prevention and Control for their advice and technical assistance on the methodological and data problems of measuring mammography performance and cancer outcomes.

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. VISA and MasterCard credit cards are accepted, also. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 37050
Washington, DC 20013

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000 or by using fax number (202) 512-6061, or TDD (202) 512-2537.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (202) 512-6000 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

For information on how to access GAO reports on the INTERNET, send an e-mail message with "info" in the body to:

info@www.gao.gov

or visit GAO's World Wide Web Home Page at:

http://www.gao.gov

United States
General Accounting Office
Washington, D.C. 20548-0001

Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100

Official Business
Penalty for Private Use $300

Address Correction Requested