Review Article

The Mammography Audit: A Primer for the Mammography Quality Standards Act (MQSA)

Michael N. Linver[1], Janet Rose Osuch[2], R. James Brenner[3], Robert A. Smith[4]

The medical audit of a mammography practice is a recognized method for evaluating mammography and the accuracy of mammographic interpretation [1–4]. As such, portions of the audit will become integral to the quality assurance activities of every mammography practice under the Mammography Quality Standards Act (MQSA) of 1992, administered by the Food and Drug Administration (FDA). The FDA Interim Rules, which became effective October 1, 1994, state that "each facility shall establish a system for reviewing outcome data from all mammography performed, including follow-up on the disposition of positive mammograms and correlation of surgical biopsy results with mammogram reports" [5]. It is expected that the proposed final rules, due to be released for public comment in 1995, will require collection of additional data for medical audits (public meeting of the National Mammography Advisory Committee, May 3, 1994). Although most mammography practices are now collecting clinical outcomes data on abnormal mammographic examinations, very few have established an organized and deliberate system of data collection necessary for a more complete mammography audit [6]. A detailed discussion of and recommendations for such an audit were recently published as part of the Quality Determinants of Mammography Guideline by the Agency for Healthcare Policy and Research (AHCPR) [7]. As members and consultants on the multidisciplinary panel that produced the guideline, we offer the following review of the various elements, definitions, and processes of the mammography audit. This is intended as a primer for all radiologists who will be performing some of the same audit activities for the MQSA.

The Mammography Audit—Its Value

In addition to meeting requirements legislated by the MQSA, the mammography audit can serve other valuable functions. First, it measures the mammographer's success in finding cancers, especially impalpable cancers, as compared with emerging national trends and goals [2–4, 8]. Regular review of individual and group audit data serves as a teaching tool, providing comparisons of performance and improving future outcomes [3]. Audit data can identify false-negative studies for review to determine their causes, allowing technical and interpretive shortcomings to be corrected [4, 8–11]. The audit can provide data for outcomes analysis locally and nationally [3, 4, 12, 13]. Audit results could improve compliance of both referring physicians and patients with screening guidelines by increasing confidence in the screening system [3, 8]. The audit is a source of data for calculating costs per patient screened, which is valuable information to radiologists preparing for capitation contracts with health care organizations [1]. Audit data can also assist in situations requiring medicolegal defense by providing a documented profile demonstrating the radiologist's ability to evaluate benign and malignant disease meeting national goals and by providing prior reference cases similar to one in contention, which substantiate the rationale for a given interpretation [3, 4, 14, 15].

The Audit Process—An Overview

The audit involves collecting and analyzing a variety of data generated from both the mammography report and any subsequent breast biopsy. The mammography report consists of demographic information, results, and recommendations, which must be constructed in forms that allow collection of useful audit data. Demographic information such as the patient's name and age requires no special coding. Results and recommendations,

Received December 28, 1994; accepted after revision February 23, 1995.
Supported by the Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services.
[1]X-Ray Associates of New Mexico, P.C., 715 Grand Ave., N.E., Ste. 112, Albuquerque, NM 87102. Address correspondence to M. N. Linver.
[2]Department of Surgery, Michigan State University, 701 N. Logan, Ste. 210, Lansing, MI 48915.
[3]The Joyce Eisenberg Keefer Breast Center, John Wayne Cancer Institute, Saint John's Hospital and Health Center, 1328 22nd St., Santa Monica, CA 90404.
[4]American Cancer Society, 1599 Clifton Rd., N.E., Atlanta, GA 30329.

however, must be categorized using standardized codes such as those of the American College of Radiology Lexicon [16] (Tables 1 and 2). This coding process establishes a standard language for data entry, facilitating data analysis [16].

Appropriately coded report information can be collated either manually or through computer software programs designed to meet the needs of mammography facilities [16, 17]. Breast biopsy results can then be acquired and coded using standard pathology nomenclature [1–4, 16, 17] (Tables 1 and 2). Integration of the mammography and pathology results then generates the important items of the mammography audit. Audit data should then be summarized and evaluated at least yearly [1–4].

The Audit Data—What to Collect? What to Calculate?

Once a data collection system with proper coding of data elements is in place, one must decide what data are essential to measure the quality of one's practice. Data collected should address the three major goals of screening mammography [18]:

1. The mammographer should find a high percentage of the cancers that exist in a given population. This percentage can be measured with cancer detection rate and sensitivity.

2. The request rates for further imaging evaluation and for biopsy should be in an acceptable range for that population. These rates can be measured with recall rate and positive predictive value (PPV).

3. Most mammographically detected cancers should have characteristics consistent with a favorable prognosis. This can be assessed by calculating the rate of minimal and node-positive cancers found mammographically.

Table 1 lists the essential raw and derived data necessary to demonstrate achievement of these goals, with one exception, which will be discussed later. Raw data refer to specific items of information, interpretive results and recommendations, and pathology findings collected directly from the mammography and pathology reports. Essential raw data include audit period dates, number of screening mammographic examinations and number of diagnostic mammographic examinations performed (see appendix for definitions), number of recalls requested, number of recommendations for surgical biopsy, biopsy results, and tumor staging.

Derived data refer to calculated measures of various mammographic and pathologic parameters based on the collected raw data. Essential derived data include number of true-positives, number of false-positives, PPV, cancer detection rate, per-

TABLE 1: The Essential Mammography Audit: The Minimum Desired Raw and Derived Data

A. Raw Data
   1. Dates of audit period and total number of examinations in that period
   2. Number of screening examinations; number of diagnostic examinations[a]
   3. Number of recommendations for futher imaging evaluation (recalls) (American College of Radiology [ACR] Lexicon Category 0 = "Needs Further Evaluation")
   4. Number of recommendations for biopsy or surgical consulation (ACR Lexicon Categories 4 and 5 = "Suspicious Findings" and "Highly Suggestive of Malignancy")
   5. Biopsy results: malignant or benign (keep separate data for fine-needle aspiration or core biopsy cases)
   6. Tumor staging: histologic type (ductal [in situ or invasive] or lobular [invasive only]), size, nodual status, and grade[b]

B. Derived data (calculated from the raw data)
   1. True-positives (TP)
   2. False-positives = three subdefinitions: FP_1, FP_2, FP_3 (see text)
   3. Positive predicitve value (PPV)
      a. If a screening/diagnostic facility, PPV can be defined any of three ways:
         1. Based on abnormal findings at screening examination (PPV_1)
         2. Based on recommendation for biopsy or surgical consultation (PPV_2)
         3. Based on result of biopsy (PPV_3, or positive biopsy rate)
      b. If a screening facility exclusively, can define only one way:
         1. Based on abnormal findings at screening examination (PPV_1)
   4. Cancer detection rate for asymptomatic (screening) cases
   5. Percentage of minimal cancers[c] found
   6. Percentage of node-positive cancers found
   7. Recall rate

[a] Separate audit statistics should be maintained for asymptomatic and symptomatic patients.

[b] The grading of tumors, although not performed as part of tumor staging by all pathologists, is nonetheless valuable information and should be collected, if available.

[c] Minimal cancer: invasive cancer ≤1 cm, or in situ ductal cancer.

TABLE 2: The More Complete Mammography Audit: Raw Data to Be Collected

1. Dates of audit period and total number of examinations in that period (usually a 12-month period).
2. Risk factors:
   a. Patient's age at the time of the examination
   b. Breast cancer history: personal or family (especially premenopausal cancer in first-degree relative—mother, sister, or daughter)
   c. Hormone replacement therapy
   d. Previous biopsy-proved atypia or lobular carcinoma in situ
3. Number and type of mammograms: screening (asymptomatic) or diagnostic (evaluation of symptoms or signs of breast cancer)[a]
4. First-time examination or routine follow-up (repeat) examination
5. Mammographic interpretation and recommendation (try to conform to American College of Radiology [ACR] Lexicon):
   a. Further imaging evaluation (recall) [ACR Lexicon Category 0 = "Needs Further Evaluation"]
   b. Routine follow-up (ACR Lexicon Categories 1 and 2 = "Negative" and "Benign Findings")
   c. Early follow-up (ACR Lexicon Category 3 = "Short-Term Follow-Up")
   d. Biopsy or surgical consultation (ACR Lexicon Categories 4 and 5 = "Suspicious Findings" and "Highly Suggestive of Malignancy")
6. Biopsy results
   a. Benign or malignant (keep separate data for fine-needle aspiration or core biopsy cases)
7. Cancer data
   a. Mammographic findings: mass, calcifications, indirect signs of malignant tumor, no mammographic signs of malignant tumor
   b. Palpable or impalpable tumor
   c. Tumor staging (pathologic): histologic type, size, nodal status, and grade[b]

Note.—Bold type indicates data desired for the essential mammography audit.

[a] Separate audit statistics should be maintained for asymptomatic and symptomatic patients.

[b] The grading of tumors, although not performed as part of tumor staging by all pathologists, is nonetheless valuable information and should be collected, if available.

centage of minimal cancers found, percentage of node-positive cancers found, and recall rate. These terms are defined below.

When the proposed final rules of the MQSA are issued, the collection of mammography data and calculation of survey statistics required by the MQSA will most likely be drawn from items listed in Table 1 (Public meeting of the National Mammography Advisory Committee, May 3, 1994).

Additional raw data for collection are listed as part of the complete raw data list in Table 2. Although not required to calculate the essential derived data of Table 1, they do provide other important information affecting audit results. For example, the ratio of first-time mammographic examinations to repeat examinations performed in a given practice can dramatically alter the rate of cancers detected overall, because the rate of cancer detection on first-time examinations is higher than that on repeat examinations [3, 19].

Additional derived data of importance can also be calculated, as listed as part of the complete derived data in Table 3. However, cost and time constraints and lack of availability of certain raw data may prohibit their calculation.

Calculation of the derived data in Table 1 or Table 3 requires categorizing every mammographic examination into one of four groups according to the following definitions, based on major audit studies in the scientific literature:

1. True-positive (TP): cancer diagnosed within 1 year after biopsy recommendation based on mammographic examination with abnormal findings [19].

2. True-negative (TN): no known cancer detected within 1 year of mammographic examination with normal findings [19].

3. False-negative (FN): detection of cancer within 1 year of a mammographic examination with normal findings [1, 2, 10, 19–22]. Although FN studies have been variably defined, this definition is the most often applied (see Appendix).

4. False-positive (FP): Three separate definitions have been used in published reports:

a. No known cancer diagnosis within 1 year of a screening mammographic examination with abnormal findings (i.e., a screening mammographic examination for which recall for further imaging evaluation or for which biopsy is initially recommended) (FP_1) [1–3, 20, 21].

b. No known cancer diagnosis within 1 year after recommendation for biopsy or surgical consultation on the basis of a mammographic examination with abnormal findings (FP_2) [1, 19].

c. Benign findings at biopsy within 1 year after recommendation for biopsy or surgical consultation on the basis of a mammographic examination with abnormal findings (FP_3) [3, 19, 20]. This definition must be distinguished from that for FP_2, because biopsy results may be unknown, or a biopsy may not always be done even when recommended in the mammographic report.

Another way to conceptualize the relationship among these four groups is expressed graphically in Figure 1 [23]. Women screened for breast cancer with mammography are placed either in the top (positive) group, if the test (i.e., the mammographic examination) indicates a suspicion of breast cancer, or the bottom (negative) group, if the test results are thought to be normal. Each group is then subdivided based on whether patients are subsequently found on biopsy to have breast cancer (left-hand columns) or not (right-hand columns). Four possible combinations then exist: if both test and biopsy are positive for cancer, this outcome is designated a TP. If both are negative for breast cancer, or if the test is negative and there is no clinical evidence of breast cancer in the absence of a biopsy, this outcome is designated a TN. If the test is positive and the biopsy is negative, this outcome is designated an FP. Conversely, if the test is negative and the biopsy positive, this outcome is designated an FN.

Given the above definitions and raw data, it is possible to now calculate the following derived data, based on major audit studies that have been published:

Sensitivity: Defined as the probability of detecting a cancer when a cancer exists, or otherwise defined as the percentage of all patients found to have breast cancer within 1 year of screening, correctly diagnosed as suggestive of breast cancer on the basis of mammographic findings [2, 3, 8, 9, 19–21, 24–26].

$$\text{Sensitivity} = TP/(TP + FN)$$

PPV: Three separate definitions may be applied, based on the above three definitions of FP:

1. PPV_1 (abnormal findings at screening): The percentage of all screening examinations with abnormal findings (i.e., those for which recall for further imaging evaluation or biopsy was initially recommended) that result in a diagnosis of cancer [2, 3, 21, 22, 24].

$$PPV_1 = TP/(\text{number of screening examinations with abnormal findings}), \text{ or } TP/(TP + FP_1)$$

TABLE 3: The More Complete Mammography Audit: Derived Data to Be Calculated

1. **True-positives, false-positives (three subdefinitions: FP1, FP2, FP3),** true-negatives, false-negatives
2. Sensitivity
3. **Positive predictive value (PPV)**
   - **a. Based on abnormal findings at screening examination (PPV1)**
   - **b. Based on recommendation for biopsy or surgical consultation (PPV_2)**
   - **c. Based on results of biopsy (PPV_3)**
4. Specificity
5. Cancer detection rate
   - **a. Cancer detection rate for asymptomatic (screening) cases**
   - b. Prevalent versus incident
   - c. Overall
   - d. Rates within various age groups
6. **Percentage of minimal cancers[a] found**
7. **Percentage of node-positive cancers found**
8. **Recall rate**

Note.—Bold type indicates data desired for the essential mammography audit.
[a]Minimal cancer: invasive cancer ≤1 cm, or in situ ductal cancer.

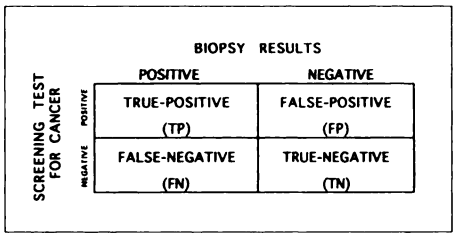

Fig. 1.—Graphic representation of relationship among true-positives (TP), false-positives (FP), false-negatives (FN), and true-negatives (TN).

2. PPV_2 (biopsy recommended): The percentage of all cases recommended for biopsy or surgical consultation (as a result of abnormal screening or diagnostic examination or additional imaging evaluation of an abnormal screening examination) that resulted in the diagnosis of cancer [1, 19].

$$PPV_2 = TP/\text{number of cases recommended for biopsy after abnormal findings on screening or diagnostic examination, or } TP/(TP + FP_2)$$

3. PPV_3 (biopsy performed): Because biopsy results may be unknown or a biopsy may not always be done even when recommended in the mammographic report, PPV_2 must be distinguished from PPV_3, which is defined as the percentage of all known biopsies done (as a result of abnormal screening or diagnostic examination or additional imaging evaluation of an abnormal screening examination) that resulted in the diagnosis of cancer. PPV_3 is also known as the biopsy yield of malignancy, or the positive biopsy rate [3, 19, 22, 24].

$$PPV_3 = TP/(\text{number of biopsies}), \text{ or } PPV_3 = TP/(TP + FP_3)$$

It is important to know which definition of PPV is being used to accurately interpret and compare audit data from a particular mammography practice with published data. For practices that do screening mammography exclusively, only PPV_1 will be of value in evaluating their data, as they will not be performing either diagnostic examinations or the further mammographic evaluation required of abnormal screening examinations for recommendation for biopsy in most cases. For practices that do both screening and diagnosis, all three definitions of PPV have value and can be applied.

Specificity: Defined as the probability of normal mammographic findings when no cancer exists, or otherwise defined as the percentage of all patients with no evidence of breast cancer within 1 year of screening, correctly identified as normal at the time of mammographic screening [1, 2, 3, 20, 21, 24, 25].

$$\text{Specificity} = TN/(FP + TN)$$

Some variation in the range of specificity will exist, depending on the definition of FP being applied, but the variation will be small because of the relatively small number of FPs and the very large number of TNs in most audit series. Because the range of variation in specificity is small, its value as a measure of mammographic interpretive quality is limited [27]. The value of specificity is further diluted by the imperfect nature of the definition of TNs: these cases are not biopsy-proved and may reveal cancer more than 1 year after a mammographic examination with normal findings.

Overall cancer detection rate: Defined as the overall number of cancers detected per 1000 patients examined by mammography [1–4, 10, 19–22, 24, 25, 27].

The cancer detection rate in asymptomatic women (see Appendix) is of greater value in the audit, as this group more closely represents the true screening population [1–3, 19].

If the appropriate raw data are available, detection rates for prevalent versus incident cancers (cancers in first-time versus follow-up mammographic examinations) [1, 3, 19, 24] and for cancers in various age groups [3, 9, 24] should also be calculated, as these provide additional valuable information.

Analyzing the Data—What Do the Numbers Tell You?

The value of calculating the derived data is in defining a mammographer's performance quantitatively. Therefore, by calculating in concert the essential data elements for providing a performance overview (cancer detection rate, sensitivity [if measurable], PPV, recall rate, tumor size, and node positivity), a mammography practice will realize benefits from a basic audit.

Desirable numerical goals toward which the mammographer should strive are listed in Table 4. These are based on a review of all major audits reported in scientific publications, as follows:

Sensitivity: The sensitivity in most recently published mammography audits is greater than 85%, using the definition given in the above section [1–4, 10, 19–21, 28]. This range is therefore thought to be a desirable goal for which to strive (Table 4).

Sensitivity may vary by age group, appearing to decrease in younger women with denser breast tissue [24]. Sensitivity is a difficult rate to calculate, requiring knowledge of the actual number of FN studies to be determined accurately (see preceding section). It is usually necessary to establish a direct link with a complete tumor registry to find the actual number of FNs [2, 10, 21, 24]. Because such a link rarely exists at this time, calculation of sensitivity is not possible for most mammography practices. Consequently, sensitivity is not considered essential to the routine audit. However, it is still useful to approximate sensitivity based on any known FN cases [3].

PPV: This number is almost always measurable, using one or more of the definitions just described. As shown earlier, published definitions vary considerably, but the most often cited is the PPV for all cases recommended for biopsy, PPV_2. A recent survey of mammography facilities showed the average PPV_2 nationally to be 21% [29]. However, a range of greater than 25% and less than 40% has been found in most recent reported series (Table 4) [3, 19, 28]. Therefore, this range should be considered an achievable goal, although most mammography practices currently do not meet this goal.

If a facility performs screening mammography exclusively, then the PPV based on the number of screening examinations with abnormal findings (PPV_1) should be used instead. This number is greater than 5% and less than 10% in most reported series [2, 3, 21, 24, 25] and should be achievable in most prac-

TABLE 4: Analysis of Medical Audit Data—Desirable Goals

Audit Data	Goal
Positive predictive value (PPV) based on abnormal findings at screening examination (PPV_1)[a]	5–10%
PPV when biopsy or surgical consultation recommended (PPV_2)	25–40%
Tumors found—stage 0 or 1[a]	>50%
Tumors found—minimal[b] cancer[a]	>30%
Node positivity[a]	<25%
Cancers found/1000 cases[a]	2–10
Prevalent cancers found/1000 first-time examinations[a]	6–10
Incident cancers found/1000 follow-up examinations[a]	2–4
Recall rate[a]	≤10%
Sensitivity (if measurable)	>85%
Specificity (if measurable)	>90%

[a]Screening cases only.

[b]Minimal cancer: invasive cancer ≤1 cm, or in situ ductal cancer.

tices. Facilities that do both screening and diagnostic mammographic examinations will also find calculation of PPV_1 of value.

If core or fine-needle aspiration biopsy is recommended, separate PPV statistics for these cases should be maintained.

PPV will vary from one practice setting to another, because of differences in patient age distribution, percentage of palpable cancers, cancer detection rate, the size and node positivity of cancers found, and the sensitivity (if measurable) [27, 30–32]. PPV is directly proportional to the age of the population being screened [3, 31]. The older the screened population, the higher the PPV will be, because there are more existing cancers in an older population.

PPV will vary directly with the size of tumors found in a screening mammography program: when most tumors being found are large, PPV tends to be higher; finding a greater percentage of small tumors usually results in a lower PPV [31].

Tumor size: In most reported series, more than 50% of tumors diagnosed by mammography are stage 0 or 1 [2, 4, 24]. More important, greater than 30% of cancers diagnosed by mammography are minimal cancers (i.e., invasive cancer ≤1 cm, or in situ ductal cancer) [1, 3, 10, 19, 21, 22, 24, 28]. Because mortality from breast cancer is directly related to tumor size [18], these percentages of small tumors found by mammography should be considered desirable goals (Table 4). Moreover, because these percentages depend on previously mentioned population factors, as well as patient compliance with screening guidelines, these numerical targets might even be regarded as minimal goals. By reaching and exceeding them, patients' outcomes are affected the most.

Tumor size will vary with the percentage of screening and diagnostic examinations in a mammography practice; symptomatic patients invariably yield larger tumors than those in a screening population [4, 24].

Node positivity: Tumor size should also be correlated with node positivity, which in most series is less than 25% in a screened population [1, 3, 4, 19, 21, 22, 24, 28]. Because mortality from breast cancer is related to the prevalence and extent of nodal metastasis, a node positivity rate of less than 25% is also a desirable goal (Table 4).

Cancers found per 1000 patients screened (cancer detection rate): This number is quite variable, with rates of two to 10 cancers per 1000 women reported in most screening series [1–4, 19–22, 24, 27] (Table 4). Variability is due to differing rates of detection in first-time screened versus already-screened patients [i.e., prevalent versus incident cancers]: prevalent cancer rates vary from six to 10 per 1000 women screened, and incident cancer rates vary from two to four per 1000 women screened [16, 19, 24] (Table 4). The cancer detection rate will also vary between younger and older populations [3, 19, 20, 24, 33]. Nonetheless, the cancer detection rate still serves as a useful measure of the effectiveness of screening mammography. For example, if an audit shows that sensitivity and PPV are both within expectations, but the number of cancers found is less than two per 1000 asymptomatic patients, then the sensitivity figure should be considered suspect. The number of cancers eluding detection in such a population is most likely too high, and the overall quality of the mammography program should be further evaluated [27, 32].

Recall rate: The percentage of patients undergoing screening mammographic examinations who are recommended for further imaging evaluation (coned compression views, magnification views, sonography, etc.) should be assessed for two reasons. First, this rate can be used to calculate one of the definitions of FP (FP_1) and one of the definitions of PPV (PPV_1) (see Derived Data section), both of specific relevance to screening mammography practices. Second, the cost-effectiveness and credibility of mammography can be negatively affected if the recall rate is disproportionately high [1]. Based on most large reported series, the percentage of patients in the screening group who are recalled for further imaging evaluation should be 10% or less (Table 4) [1–3, 19, 21, 24]. Many authors have also noted that this rate may decrease with increasing experience [1, 3].

Specificity: Specificity is usually found to be greater than 90% [2, 21, 24] (Table 4). However, it is not even calculated in many large studies [1, 3, 4, 19], as its calculation requires knowledge of all TNs, a number which in turn is based on the number of FNs. The number of FNs is usually the least accessible data in any audit. For this reason, and for those cited previously, specificity is not considered essential to a routine audit.

Further Benefits: The Audit as a Teaching Tool

The audit has significance as a teaching tool regarding three other specific issues. First, a group audit may be reviewed in tandem with individual audits. Pooling the data of all individuals within a group gives greater statistical power to audit results, facilitating comparison to expected results such as those in Table 4 [3, 19]. However, the multiple variables described earlier (prevalent versus incident cancers, age of a population, ratio of screening to diagnostic mammograms, etc.) that markedly influence group audit results may render comparisons to other group audits less valuable than an intragroup audit of individuals.

A major advantage to an individual audit is in providing a valid objective comparison among group members. If certain group members show considerable variance from others when performance standards are compared, measures can be taken to improve the performance of those at variance and thus improve future outcomes [3, 19].

The second issue concerns the review of FNs. As mentioned earlier, these cases may be difficult to identify if access to a complete tumor registry is not possible [3]. However, if available for review, all FN cases should be evaluated thoroughly to assess cause (technical versus interpretive error) [4, 9–11]. Their real value is educational: by critically reviewing such cases, a group can benefit all its members by improving overall quality and, in turn, future outcomes. Group review of all interval cancers, regardless of the interval between the last mammographic examination interpreted as normal and the detection of cancer, can also be of value for the same reasons [9, 10].

The third issue is one that many practices are already addressing: review and comparison of pathology reports of breast biopsies with the corresponding mammographic examinations that prompted those biopsies. Correlations between mammographic and histologic findings in cases of both malignant and benign pathology have immeasurable teaching value. Review of cases by the mammographer and the pathologist together can further enhance the learning process for both individuals.

Sources for Audit Data

As stated previously, patients' demographic information and pertinent mammographic results and recommendations

should be available from a well-designed and properly coded mammography report record, especially if it is computerized [3, 17, 19, 24]. Biopsy results are available from a variety of sources [1–4, 19, 21, 22, 24]. Malignant biopsy results can be found through a complete regional or statewide tumor registry. If linkage to a tumor registry is anticipated, additional patient identifiers may be needed to match mammography and registry data. If such a registry does not exist or access to its data is not possible, definitive diagnosis of cancer can be obtained from, in order of preference, the pathology report, the referring physician or surgeon, or the patient herself. Benign biopsy results will not be collected by most tumor registries and must be obtained from the alternative sources above.

The importance of attempting to obtain complete follow-up on every patient with suspicious findings should be stressed. Published audit results have shown that it is not possible to obtain complete follow-up on every patient with suspicious mammographic findings, even when linkage to a tumor registry is established [2, 19]. Nonetheless, efforts to obtain this follow-up information should include the methods just described.

Medicolegal Considerations

At this time, all states have statutes in place that protect from discovery peer review records generated by a structured peer review committee in the hospital setting [4, 34]. However, virtually no statutes exist to protect from discovery all other information generated in the hospital under the auspices of organized quality review activities, or information reported outside the peer review setting, or any quality review information in the outpatient setting [35].

Therefore, at this time, it is more appropriate that complete mammography audits be maintained primarily as internal audits. Interpreting physicians should not disseminate the audit data more widely without being aware of confidentiality legislation in their state and the waiving of limited peer review privilege.

Model legislation does exist: Congress provided protection to participants of quality control programs and created a qualified immunity for the medical quality assurance records generated by the programs within the military health care system (10 USC 1102) and the Department of Veteran Affairs (38 USC 5705). However, such broadly drawn protective legislation does not otherwise exist at this time. Consequently, the issues of discoverability of audit data and the relationship to the MQSA legislation are currently under active review by the MQSA National Mammography Quality Assurance Advisory Committee and the FDA (Public meeting of the National Mammography Advisory Committee, May 3, 1994).

Summary

The mammography medical audit is a recognized measure of the interpretive ability of the mammographer and a means of quantifying the success of mammography in detecting early breast cancer. Because it is a significant component of mammography quality assurance, some form of the audit will most likely be included in the MQSA.

Once a data collection system with proper coding of data elements is in place, then collection, calculation, and analysis of appropriate raw and derived data for either a basic or a more complete audit should be done at least yearly and should answer the three essential questions that determine a mammographer's success: (1) Are the cancers that exist being found? (2) Are these cancers being found with an acceptable number of recalls and biopsies? (3) Are a large proportion of these cancers small and node-negative? By answering these questions with quantitative data, it is possible to compare the mammographer's performance to the range of desirable values found in other audits reported throughout medical publications and to prior performance.

Additional audit activities such as evaluating group audit versus individual audit statistics, reviewing FN and other interval cancer cases, and correlating pathology reports with the corresponding mammographic examinations are teaching tools that result in improved clinical outcomes.

Legal constraints on the discoverability of mammography data may deter implementation of optimal audit programs and are currently being addressed by legislative efforts.

In sum, the audit process required by the MQSA and outlined as above offers radiologists the opportunity to add a greater measure of quality to their mammography practices and, more important, to the lives of the patients they serve.

ACKNOWLEDGMENTS

We thank the following fellow members and consultants on the AHCPR Panel on Quality Determinants of Mammography who also helped develop the section of the Clinical Practice Guideline on the medical audit: Carl J. D'Orsi, Lawrence W. Bassett, R. Edward Hendrick, Victor Hasselblad, and Darryl Carter. For their help in preparing the manuscript, we thank Robert D. Rosenberg, Richard E. Bird, Peter J. Dempsey, and Charles A. Kelsey. We also extend special thanks to AHCPR staff member Marietta Anthony, whose efforts brought the guideline to timely fruition.

APPENDIX

Screening and diagnostic mammographic examinations: A screening examination is one performed on an asymptomatic woman to detect early, clinically unsuspected breast cancer. There are two distinct types of diagnostic mammographic examinations. The first is that performed on a woman with clinical signs or symptoms that suggest breast cancer and, for purposes of the audit, is the only one considered a diagnostic mammographic examination. The second type is that performed on a woman for whom further mammographic evaluation has been requested because of an abnormal screening mammographic examination. For audit purposes, the mammographic and consequent pathologic findings in the latter type should be included with the data collected and calculated for the screening population, because the evaluation was initiated by a screening mammographic examination. (Two other special screening examinations, that performed in a woman with a history of breast cancer with breast conservation and that performed in a woman with augmented breasts, are often defined as diagnostic but for audit purposes should be included in the screening group.)

False-negative (FN): This term has been defined many ways throughout the literature. The 1-year definition in the text is best suited for audit purposes because it allows valid, consistent, and timely comparisons to be made for an individual or a group. In addition, the 1-year time frame is matched to the preferred screening interval for the largest number of women screened, those over age 50, and is well within the bounds of the estimated average lead-time [18]. Further, it is the definition quantified most completely in reported audit studies [2, 10, 19, 24]. Accordingly, it is the definition from which a consistent range of values for sensitivity (calculated from the FN numbers) has been historically established and that can be used as a standard to evaluate mammography data.

Many other definitions of FN exist, each with merit. These include (1) Any palpable or impalpable cancer detected subsequent to a mammographic examination interpreted as normal, regardless of the length of

time between that mammographic examination and the moment of detection [1, 3, 21, 36]. Each cancer included under this definition should be reviewed by the mammographer for its teaching value as a missed cancer, as recommended in the section on the audit as a teaching tool. However, the open-ended nature of this definition of FN renders it impractical for audits designed to measure data over finite and relatively brief periods. (2) Any cancer detected within 4 months (or, in some series, 6 months) of a normal mammographic examination [28]. This definition is considered to be too limited in its scope and is also not matched to the ideal 1-year screening interval that applies to most women. (3) Any palpable cancer detected between a normal screening mammographic examination and the expected time of the next routine screening examination [18, 22, 28]. This definition functions well as a measure of the success of mammography within given screening intervals. However, it does not evaluate impalpable interval cancers and may have less instructive value.

All FN definitions are fraught with problems. One dilemma is whether only those missed cancers that are visible in retrospective review on previous mammographic examinations should be considered FNs. This view involves a consideration of threshold and subthreshold features of malignancy [36]. Ideally, a blind review by one or more radiologists should be done to provide an unbiased evaluation of such cases, but even under these conditions, one has the unavoidable ability to see a cancer on the prior examination when the cancer is known to exist on the present study. Another dilemma is encountered when double reading of mammographic examinations is done and only one of the two readers correctly identifies the cancer. A problem unique to the 1-year definition is the situation in which a woman returns for screening less than 1 year since her last screening study and in whom a cancer is now found. The cancer is considered an FN by this definition, but because it has been found on the next routine screening examination, it may be viewed as a TP.

None of these problems will be universally resolved. However, for the purposes of comparing a mammographer's audit data from one yearly audit period to the next, and further comparing yearly data from one mammographer or practice to the next, both the 1-year definition of the FN and the derived definition of sensitivity as described in the text remain the most objective and widely used at this time.

Asymptomatic and symptomatic women: Asymptomatic women are defined as those presenting for screening mammographic examinations with no known signs or symptoms of breast cancer at the time of their examinations. The authors include in this group women who may not have had physical examinations prior to their mammographic examinations or in whom lesions are palpated in retrospect, as the women in both these subgroups are part of the screening pool at the time they present for their mammographic examination.

Symptomatic women are those who present for mammographic examination because of symptoms or signs of possible breast cancer. Included in this group are women referred for evaluation because of abnormal breast physical examinations by their clinicians but in whom screening mammograms are performed because the clinicians never related information about the abnormal physical findings to the mammographer. Because women in this subgroup are not part of the screening pool at the time they present for mammographic examination, they should be placed with the symptomatic group for audit purposes. Even if this subgroup is included in the asymptomatic category, audit statistics will not be changed for the vast majority of facilities, as this subgroup is small compared with all asymptomatic women screened by mammographic examination.

REFERENCES

1. Bird RE. Low-cost screening mammography: report on finances and review of 21,716 cases. *Radiology* **1989**;171:87–90
2. Robertson CL. A private breast imaging practice: medical audit of 25,788 screening and 1,077 diagnostic exams. *Radiology* **1993**;187:75–79
3. Sickles EA. Quality assurance: how to audit your own mammography practice. *Radiol Clin North Am* **1992**;30:265–275
4. Spring DB, Kimbrell-Wilmot K. Evaluating the success of mammography at the local level: how to conduct an audit of your practice. *Radiol Clin North Am* **1987**;25:983–992
5. MQSA Interim Rules. Federal Register. Washington, DC: Government Printing Office, December 21, 1993;58:67558–67572
6. Brown ML, Houn F. Quality assurance audits of community screening mammography practices: availability of active follow-up for data collection and outcome assessment. *AJR* **1994**;163:825–829
7. Bassett LW, Hendrick RE, Bassford TL, et al. Quality determinants of mammography. Clinical Practice Guideline No. 13. AHCPR Publication No. 95-0632. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, U.S. Department of Health and Human Services, October 1994
8. Sickles EA, Ominsky SH, Sollitto RA, et al. Medical audit of a rapid-throughput mammography screening practice: methodology and results of 27,114 examinations. *Radiology* **1990**;175:323–327
9. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology* **1992**;184:613–617
10. Burhenne HJ, Burhenne LW, Goldberg F, et al. Interval breast cancers in the Screening Mammography Program of British Columbia: analysis and classification. *AJR* **1994**;162:1067–1071
11. Moskowitz M. Interval cancers and screening for breast cancer in British Columbia (commentary). *AJR* **1994**;162:1072–1075
12. Clark RA, King PS, Worden JK. Mammography registry: considerations and options. *Radiology* **1989**;171:91–93
13. Hurley SF. Screening: the need for a population register. *Med J Australia* **1989**;153:310–311
14. Brenner RJ. Medicolegal aspects of breast imaging. *Radiol Clin North Am* **1992**;30:277–286
15. Reinig JW, Strait CJ. Professional mammographic quality assessment program for a community hospital. *Radiology* **1991**;180:393–396
16. Kopans DB, D'Orsi CJ, Adler DED, et al. Breast imaging reporting and data system. Reston, VA: American College of Radiology;**1993**
17. Sickles EA. The use of computers in mammography screening. *Radiol Clin North Am* **1987**;25:1015–1030
18. Tabar L, Fagerberg G, Duffy SW, et al. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* **1992**;30:187–210
19. Linver MN, Paster SB, Rosenberg RD, et al. Improvement in mammography interpretation skills in a community radiology practice after dedicated teaching courses: 2-year medical audit of 38,633 cases. *Radiology* **1992**; 184:39–43
20. Braman DM, Williams HD. ACR accredited suburban mammography center: three year results. *J Fla Med Assoc* **1989**;76:1031–1034
21. Lynde JL. A community program of low-cost screening mammography: the results of 21,141 consecutive examinations. *South Med J* **1993**;86: 338–343
22. Moseson D. Audit of mammography in a community setting. *Am J Surg* **1992**;163:544–546
23. D'Orsi CJ. Screening mammography pits cost against quality. *Diagn Imaging* **1994**;16:73–76
24. Burhenne LJW, Hislop TG, Burhenne HJ. The British Columbia mammography screening program: evaluation of the first 15 months. *AJR* **1992**;158:45–49
25. Baines CJ, Miller AB, Wall C, et al. Sensitivity and specificity of first screen mammography in the Canadian National Breast Screening Study: a preliminary report from five centers. *Radiology* **1986**;160:295–298
26. Schmidt RA, Metz CE. Please be specific (letter). *AJR* **1990**;154:1121–1122
27. Ciatto S, Cataliotti L, Distante V. Nonpalpable lesions detected with mammography: review of 512 consecutive cases. *Radiology* **1987**;165:99–102
28. Margolin FR, Lagios MD. Development of mammography and breast services in a community hospital. *Radiol Clin North Am* **1987**;25:973–982
29. Brown ML. Personal communication, **1994**
30. D'Orsi CJ. To follow or not to follow, that is the question. *Radiology* **1992**;184:306
31. Kopans D. The positive predictive value of mammography. *AJR* **1992**;158: 521–526
32. Moskowitz M. Predictive value, sensitivity and specificity in breast cancer screening. *Radiology* **1988**;167:576–578
33. Moskowitz M. Breast cancer: age-specific growth rates and screening strategies. *Radiology* **1986**;161:37–41
34. American Medical Association. *A compendium of state peer review immunity laws.* Chicago: American Medical Association, **1988**:vi
35. American Medical Association. Report of AMA reference committee G, substitute resolution 722, *Medical peer review outside hospital settings.* Chicago: American Medical Association, **1992** June
36. Brenner RJ. Followup as an alternative to biopsy for probably benign mammographic abnormalities. *Curr Opin Radiol* **1991**;3:588–592