Measuring Mammography Compliance: Lessons Learned from a Survival Analysis of Screening Behavior

Melissa R. Partin, Amy L. Casey-Paal, Jonathan S. Slater, and Jane E. Korn
Minnesota Department of Health, Cancer Control Section, Minneapolis, Minnesota 55440-9441

Abstract
Data from a telephone survey of women participating in a federally funded screening program were used to demonstrate the sensitivity of mammography compliance estimates to varying definitions of the time interval within which women are considered compliant with screening guidelines and what constitutes a true screening (as opposed to diagnostic) mammogram. The survival analysis approach used reveals patterns concealed by other approaches to measuring mammography behavior and provides a means for quantifying the impact of various definitional choices on compliance estimates. The results suggest that, although variations in defining and excluding potential diagnostic mammograms lead to differences in compliance measures no greater than 6%, differences as small as 1 month in the screening interval definition used can produce differences in compliance estimates as large as 27%. These results call into question the comparability of estimates across studies and suggest that standard measures would greatly facilitate future efforts in understanding how to promote compliance with mammography screening guidelines.

Introduction
Breast cancer is the most common cancer affecting women in the United States and the second leading cause of cancer deaths in this population (1). Because there are no established modifiable risk factors for the disease, the best defense against breast cancer mortality is early detection coupled with effective treatment. Screening mammography, the most efficacious means of early cancer detection presently available, has been shown in clinical trials to have the potential to reduce breast cancer mortality by up to 30% (depending on the age group examined) if the majority of women receive screening mammograms every 1–2 years (2). Based on the results of these trials, many professional organizations have developed and endorsed specific mammography screening guidelines. However, the age appropriateness and periodicity of screening mammograms remain a matter of debate. For instance, the National Cancer Institute recently published a statement recommending that women age 40 and older be screened every 1–2 years (3), but the American Cancer Society now recommends annual mammograms for women beginning at age 40 (4). The American College of Obstetricians and Gynecologists, on the other hand, recommends women of ages 40–50 be screened every 1–2 years, and that women >50 be screened annually (5). The United States Preventive Services Task Force recommends mammography screening every 1–2 years for women ages 50–69 but makes no recommendations for women <50 or >69, arguing that there is insufficient evidence to recommend for or against routine screening in these age groups (6). Despite these differences, these and other organizations have consistently agreed that women ages 50–69 should receive screening mammograms every 1–2 years (3–8). Because the mortality reductions observed in clinical trials depend on sustained, routine screening, compliance with screening guidelines has become one of the most common ways to measure the success of public health efforts to promote mammography use. However, lack of comparability of estimates across studies may hamper our ability to determine which of these efforts are most successful at enhancing compliance.

A common definition of mammography compliance used in previous studies is having been screened at least once in the past 2 years for women ages 40–49 and having been screened within the past year for women age 50 and older. So defined, baseline compliance estimates published in recent studies that we reviewed range anywhere from 18 to 64% for women 40 and older (9–16) and from 15 to 26% for women 65 and older (17–19). The breadth of this range raises a number of questions. Do the differences reflect actual differences in the screening behavior of the populations examined, differences in the methodology used to define and measure compliance, or both? Because most published studies supply little detail on the methodology used to operationalize definitions of compliance, the actual extent to which methodological differences account for variation in published estimates is unknown. This study demonstrates the potential for substantial variation in compliance estimates across even subtle differences in methodology and discusses the implications of this variation.

Although several methodological factors could account for variation in compliance estimates (including study design and data collection procedures), this study focuses on the role played by two definitional issues: the definition of the time interval within which women are considered compliant with mammography guidelines; and the definition of what constitutes a true screening mammogram.

tutes a true screening (as opposed to diagnostic) mammogram. As mentioned previously, compliance with mammography guidelines is most commonly defined as screened “within the past year” (for women of age 50 and older). The few studies using this definition that have provided sufficient methodological detail to determine how the “past year” interval was actually operationalized in the analysis have used definitions that imply an interval of less than 12 months. Most of these studies defined mammography compliance as having had a mammogram in a particular calendar year (16–22), although one study asked women to indicate whether their last mammogram was <12 months ago or some other interval (23). However, for the majority of studies reporting compliance estimates, it is not clear how strictly the “past year” interval is operationalized in the analysis. Is it operationalized as screened within <12 months, ≤12 months, <13 months, or some other interval? Although the distinctions between these definitions may seem trivial, if the majority of screening activity in fact occurs approximately (but not precisely) at 12-month intervals, these subtle definitional choices may yield substantially different estimates. This analysis demonstrates the sensitivity of compliance estimates to different screening interval definitions.

Mammograms can be used for both screening and diagnostic purposes. A screening mammogram is conducted in the absence of any breast problems or symptoms for routine early detection purposes only. A diagnostic mammogram is conducted solely to form a diagnostic conclusion after an abnormal breast exam, mammogram, or presenting symptom. Because diagnostic mammograms may inflate estimates of screening mammography use, some investigators have attempted to exclude them from compliance measures. Previous studies based on self-reported mammography histories have omitted diagnostic mammograms either by limiting the study sample to women receiving referrals for screening mammograms (24) or by excluding mammograms reportedly received for purposes other than a routine exam (11, 25–28). Studies based on medical and administrative data have made these exclusions by referring either to the reason for the mammogram referral recorded on the patient record (29, 30) or the procedure code recorded on a mammogram claim (16, 31). However, most studies do not mention diagnostic exclusions, and some claim screening and diagnostic distinctions cannot be reliably made with available data (18). What are the consequences of failing to exclude diagnostic mammograms from estimates of the levels and determinants of mammography compliance? This analysis addresses this question by demonstrating the range of compliance estimates for a single population that can result from various approaches to defining (and excluding from estimates) potential diagnostic mammograms.

This study uses survival analysis techniques to demonstrate the range of compliance estimates that can result from varying definitions of screening intervals and diagnostic mammograms. This dynamic analytic approach reveals patterns of mammography use concealed by the static compliance measures used in previous studies and provides a means for quantifying the impact of various definitional choices on compliance estimates.

Materials and Methods
Sample. The data presented are from a survey of women participating in the MBCCCP. Funded by the Centers for Disease Control since 1991, the MBCCCP is part of a national effort to increase cancer screening among low income, uninsured, and underinsured women. The MBCCCP provides free mammograms and Pap smears to income-eligible women through a network of over 250 hospitals and clinics around the state. To date, the MBCCCP has screened over 53,000 women and provided over 40,000 mammograms. In the fall of 1995, a random sample of 500 MBCCCP participants were recruited to participate in a computer-assisted telephone survey on mammography behavior. Because it is possible for MBCCCP participants to obtain future mammograms from providers outside the program network, MBCCCP records provide incomplete information on the repeat mammography use of program participants. Hence, the survey was conducted to obtain more complete information on repeat mammography use than was available from program data alone. The survey sample was restricted to women age 40 and older who had received a screening mammogram through the MBCCCP between May and November of 1994. Women diagnosed with breast cancer through the MBCCCP were excluded. Women were interviewed between 13 and 18 months after having had a mammogram through the program and asked about their mammography history, their knowledge and attitudes about the efficacy of mammography, their satisfaction with the program, and their demographic characteristics. The 13- to 18-month study window was chosen on the basis of the competing objectives of allowing enough time for women to have been rescreened, and keeping recall errors, which increase with time, to a minimum. Three women were determined to be ineligible because they either presently had breast cancer, had moved out of the country, or were deceased. Interviews were completed with 424 of the 497 eligible women, resulting in an overall response rate of 85.3%. Because women in this study were followed for <24 months (the screening interval most commonly used to assess mammography compliance for women ages 40–49), the analyses presented are limited to women 50 and older (n = 288).

Measures. Women participating in the survey were asked whether they had received any mammograms since their previous mammogram through the program, and if so, how many. They were then asked to provide the exact month and year of each mammogram reported. Women providing a range or season as their first answer were asked to provide the most likely month within the range or season. Women providing uncertain answers were probed for the most likely month and year, and as a last resort, to estimate how many months ago their last mammogram occurred. One woman was deleted from the analysis because she was not able to provide a specific mammogram date even after probing, and another because she reported a repeat mammogram date that was earlier than her previous mammogram date, leaving 286 women in the sample analyzed. In addition to the questions on the number and timing of mammograms, for each mammogram reported, women were asked to indicate whether the mammogram was done as part of a routine check-up or because of some breast problem they were experiencing.

Analysis. Survival analysis techniques were used to study the occurrence and timing of repeat mammograms reported in the survey. A myriad of survival models are available to choose from, but a common feature of all of them is that they permit the analysis of right censored cases. A case is said to be right censored if all that is known about the event of interest in this case is that it has not occurred before the end of the observation period. In the present analysis, women are considered right.
censored if they do not report a repeat mammogram occurring before the date they are interviewed.

With precise information on whether and when the event of interest (in this case, a mammogram) occurred, survival models can be used to derive a number of useful summary statistics including the average waiting time to the event of interest, the rate at which the event occurs at different intervals, and the cumulative proportion of observations experiencing the event over time. Using the actuarial method of constructing survival functions available in Statistical Analysis Software, we derived estimates of both monthly and cumulative repeat screening rates by time since the previous mammogram and then used these estimates to create a picture of the natural history of repeat screening behavior in the study cohort. This natural history reveals patterns concealed by the dichotomous measures of compliance with mammography guidelines used in previous studies of screening behavior and highlights some of the implications of methodological variation in how measures of compliance are formulated.

To illustrate the range of compliance estimates that can result from various screening interval definitions, we first estimate the monthly rates of repeat screening (or the "hazard rates" of repeat screening) by time since the previous mammogram. These rates are estimated from the following formula:

\[ h(t_i) = \frac{d_i}{n_i} \frac{w_i}{d_i} \]

where, for the i-th month, \( t_i \) is the midpoint, \( d_i \) is the number of cases reporting they had repeat mammograms in that month, \( n_i \) is the number of cases who have neither had a repeat mammogram nor been censored before the beginning of the month, and \( w_i \) is the number of cases censored within the month. We plot these hazard estimates by time since the previous mammogram to provide a visual tool for assessing at what intervals women previously screened are most likely to return for a subsequent mammogram.

Variation in screening interval definitions within the range of when women are most likely to return can have a substantial impact on compliance estimates. To illustrate this, we next estimate the survival function. The survival function derived from the actuarial method is calculated as:

\[ \hat{S}(t_i) = \prod_{j=1}^{i-1} (1-q_j) \]

where \( t_i \) represents the beginning of the interval of interest and \( q_j \) the conditional probability of repeat screening or:

\[ \frac{d_i}{w_i} \frac{N_i}{N} \]

In this analysis, one minus the survival function (or the failure function) represents the cumulative proportion of women screened at various monthly intervals since the previous mammogram. This function is plotted to illustrate how compliance estimates can vary across screening interval definitions.

To illustrate the potential impact of diagnostic mammograms on compliance estimates, the failure function is recalculated three more times: the first time treating as censored women who had mammograms <6 months after their previous mammogram; the second time treating as censored women who had mammograms <10 months after their previous mammogram; and the last time treating as censored women who had mammograms for reasons other than a routine exam. The estimates resulting from these three different models were then compared with the estimates making no distinction between screening and diagnostic mammograms.

### Results

#### Sample Characteristics

Table 1 displays the demographic characteristics of the study sample. Reflecting the fact that the MBCCCP is targeted to underserved women, >99% of the sample reported incomes at or below 250% of the federal poverty level, and ~20% reported they did not have health insurance at the time of the survey. Consistent with the state of Minnesota as a whole, the majority (94%) of women in the sample were Caucasian, and <1% were Hispanic. Approximately 68% of the sample falls within the 50–64 age group, and 32% were age 65 and older.

The descriptive statistics on the number and timing of repeat mammograms reported by study participants suggest some specific patterns that might be considered more consistent with diagnostic follow-up than screening behavior. For instance, ~2% of women in the sample reported having two mammograms in the study window, indicating they had at least two mammograms in a 1-year period. Looking at the timing of and reason for the mammogram, ~7% reported having mammograms <10 months after their previous mammogram, and ~6% reported having mammograms for reasons other than a
routine exam. These patterns raise the question of how much our estimates of compliance may be affected by reported diagnostic mammograms. The sensitivity of compliance estimates to both reported diagnostic mammograms and screening interval definitions is assessed in the analyses that follow.

**Sensitivity of Compliance Estimates to Screening Interval Definitions.** Fig. 1 displays the hazard rates of repeat screening among women in our sample. In this figure, the data points correspond to the hazard rates observed at the mid-point of each month plotted on the X-axis. For instance, the first data point corresponds to the hazard rate observed at the mid-point between zero and 1 month from the previous mammogram. The figure suggests that women are most likely to return for a mammogram between 12 and 13 months after their previous mammogram. However, the hazard rates are very low just prior to 12 months but remain relatively high just after 12 months.

Fig. 2 displays the cumulative proportion of women returning for a subsequent mammogram at various monthly intervals since the previous mammogram. In this figure, the data points correspond to the cumulative proportion of women screened before the beginning of each month plotted on the X-axis. For instance, the last data point corresponds to the cumulative proportion of women screened <18 months after their previous mammogram. According to this figure, if the screening interval is defined as <12 months, only 16% of women in our sample would be considered compliant with mammography screening guidelines. By contrast, if the screening interval was defined as 12 completed months (<12 months), 43% would be considered compliant. At least one other study has used a screening interval definition of 15 completed months, arguing that this interval is more appropriate than the standard 12-month interval because it allows for
both scheduling changes and lags between mammography referral and mammography completion. If this more liberal screening interval is used, our compliance estimate would be approximately 64%.

**Sensitivity of Compliance Estimates to Definitions of Diagnostic Mammograms.** Table 2 displays estimates of the cumulative proportion of women screened at various intervals resulting from the three different definitions of what could constitute a diagnostic (as opposed to screening) mammogram investigated. These estimates are compared with those making no distinctions between screening and diagnostic mammograms (column 1). The comparisons reveal that for each definition of diagnostic mammograms used, the differences between estimates based on no restrictions and those excluding potential diagnostics are greatest for screening intervals <12 months and decrease at wider intervals. The definition excluding mammograms occurring <10 months after the previous mammogram is clearly the most conservative used and produces the greatest differences from unrestricted estimates. However, even when this conservative definition is used, adjustments for potential diagnostic mammograms decrease compliance estimates by no more than 6%.

**Discussion**

Although the proportion of women 50 years of age and older having a mammogram in the past 1–2 years is approaching the Year 2000 goal of 60% for this outcome (32), significant population reductions in breast cancer mortality will not be achieved until compliance with mammography screening guidelines is practiced by many women over extended periods of time. Because it shapes the results of surveillance, intervention development, and evaluation, how we measure this compliance has broad implications, affecting the work of researchers, clinicians, and public health policy makers.

This study quantifies the potential impact of various definitions of screening intervals and diagnostic mammograms on estimates of compliance with mammography screening guidelines. The results have many important implications. The most obvious implication of the findings on variation in estimates across screening interval definitions is that compliance estimates based on even slightly different definitions are not comparable. As stated previously, although the most commonly used measure of screening compliance for women age 50 and older is having been screened in the past year, it is seldom clear from studies reporting on such measures just how strictly the "past year" interval is defined. This study suggests that differences as small as 1 month in the screening interval definition used can lead to differences in compliance estimates as large as 27%. This lack of comparability across studies makes it difficult to assess which programs and policies are most effective in meeting common screening goals. The most serious consequence of this difficulty is that potential errors resulting from it can lead to inappropriate intervention and health policy decisions.

A less obvious implication is that too conservative a definition of the screening interval may lead to erroneous conclusions about the determinants of compliance. For instance, if compliance is narrowly defined as screened <12 months after the previous mammogram, women screened between 12 and 13 months after the previous mammogram would be labeled noncompliant. If these women are more similar to those in the compliant group than others in the noncompliant group (which seems highly likely), and they represent a sizable proportion of all women rescreened (44% in the present study), classifying them as noncompliant may attenuate estimates of the determinants of compliance, and thereby mask potentially important targets for intervention.

The results of the analyses comparing life table estimates derived from three different definitions of what constitutes a true screening mammogram mirror the results of the screening interval analysis in that they suggest that compliance estimates based on different distinctions between screening and diagnostic mammograms are not comparable. However, the variation in estimates across different definitions is likely to be small, producing differences no greater than 6%. Furthermore, because screening intervals of <12 months (most typically 6 months) are not infrequently recommended for women with a personal history of breast problems, uniformly classifying mammograms obtained at intervals <10 months as diagnostic is likely an overly conservative approach. Even conservative approaches to defining and excluding diagnostic mammograms from screening compliance measures are not likely to have a great impact on estimates, however. Hence, of the two methodological factors discussed, variation in the definition of the screening interval poses a much greater threat both to the comparability of estimates across studies and to the accurate estimation of determinants of screening compliance.

Because the data presented in this manuscript are based on self-reported mammography histories, they may suffer a number of limitations, including potential overreporting of mammograms and the tendency to round mammogram dates to 12-month intervals. However, for several reasons we believe that the self-reported mammography history data obtained from the MBCCCP mammography survey provides highly accurate estimates of repeat mammography levels and patterns:

(a) For women who returned to MBCCCP for screening mammograms in the study window, we were able to compare self-reported mammogram dates to dates recorded in their medical records. We found that the accuracy of the self-reported mammogram dates obtained in the MBCCCP mammography survey exceeded the accuracy reported in previous studies (31–37). Approximately 56% of women who returned to MBCCCP for their subsequent mammograms provided a self-reported mammogram date that matched the clinic recorded

---

**Table 2** Estimates of the cumulative proportion screened by month since the previous mammogram and type of diagnostic mammogram exclusion used

<table>
<thead>
<tr>
<th>Months since previous mammogram</th>
<th>No exclusions</th>
<th>Mammograms &lt;6 months after previous excluded</th>
<th>Mammograms &lt;10 months after previous excluded</th>
<th>Mammograms obtained for reasons other than routine exam excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cumulative proportion</td>
<td>Cumulative % decrease relative to no exclusions</td>
<td>Cumulative proportion</td>
<td>% decrease relative to no exclusions</td>
</tr>
<tr>
<td>&lt;12</td>
<td>0.1573</td>
<td>0.1268</td>
<td>0.03</td>
<td>0.094</td>
</tr>
<tr>
<td>&lt;13</td>
<td>0.428</td>
<td>0.4073</td>
<td>0.02</td>
<td>0.385</td>
</tr>
<tr>
<td>&lt;15</td>
<td>0.5912</td>
<td>0.5764</td>
<td>0.01</td>
<td>0.5605</td>
</tr>
</tbody>
</table>
mammogram date exactly with respect to month and year, 87% provided dates that matched clinic dates within 3 months, and 93% accurately reported their last mammogram within the year. Additionally, the fact that the hazard rate graphs for the self-reported (Fig. 1) and medical record-based (not shown) mammography histories of women participating in this survey are virtually identical suggests that the peaks in repeat mammogram rates at 12 months found in this analysis are real, rather than artifacts of rounding errors in self-report.

Because this survey was conducted prospectively and the date of the respondent’s previous mammogram was given to them during the interview to bound their recall, no respondent had to recall back more than 20 months when asked about their mammography history. By contrast, in retrospective surveys with no information on the previous mammogram date available to bound the respondent’s recall (for examples, see Refs. 9, 11, 13, 15, 21, 26–28, 33–35, and 37), women may have to recall anywhere from 0 days to decades when asked about their mammography history. Because the accuracy of temporal memory decreases with time (38), shorter recall periods generally produce more accurate and precise estimates than longer ones.

(c) As described earlier, we used a sophisticated set of probes for converting vague or uncertain responses to mammography history questions to more precise dates. Although of superior quality, the accuracy of mammography histories collected for this study is secondary to our primary thesis, which is that estimates of compliance with mammography screening guidelines are highly sensitive to methodological choices. Of the various methodological factors that could affect estimates of compliance, we have chosen to focus on two definitional issues that apply to both self-reported and record-based data sources. However, any future efforts to enhance the comparability of estimates across studies should also consider the impact of the study design and data collection issues not addressed in this report.

This study demonstrates how sensitive mammography compliance estimates can be to even subtle methodological choices and calls the comparability of estimates across studies into question. Without comparable estimates, progress toward understanding how to best enhance mammography compliance will be compromised. To facilitate future efforts in promoting compliance with mammography screening guidelines, we offer the following recommendations for studies estimating levels of mammography use:

(a) Because not all sources of data on mammography history permit investigators to make reliable distinctions between screening and diagnostic mammograms, and the results of this analysis suggest that such distinctions are not likely to impact compliance estimates significantly anyway, future studies should not bother to try to eliminate diagnostic mammograms from the numerator of compliance estimates.

(b) Studies citing estimates of mammography compliance should clearly specify not only how compliance was defined for the analysis but also how the data on mammography history used to classify women into compliant and noncompliant groups were collected and coded. With regard to coding, enough detail should be provided to permit the reader to distinguish between screening interval differences as subtle as groups were collected and coded. With regard to coding, of this analysis suggest that such distinctions are not likely to the following recommendations for studies estimating levels of compliance with mammography screening guidelines, we offer choices and calls the comparability of estimates across studies

(c) As long as annual mammography is the most common screening regimen recommended for women age 50 and older, surveillance studies monitoring strict compliance with this recommendation (e.g., the Behavioral Risk Factor Surveillance System and the National Health Interview Survey) will likely continue to use the imprecisely defined “screened in the past year” measure. Although we believe a more liberal definition of compliance would be preferable, as explained below, we recommend that future studies using this definition operationalize the “past year” interval as 12 completed months (or <13 months) rather than less than 12 months. This argument is based on the finding from this study that the bulk of women returning for repeat mammograms do so between 12 and 13 months after their previous mammogram. When either testing the effectiveness of interventions for promoting compliance with screening guidelines or designing recall systems for women “past due” for mammography, however, we recommend using a more liberal screening interval of 15 completed months (or less than 16 months). Our reasons for this last recommendation are as follows:

(a) As argued in previous studies, even the most diligent women who begin the process of scheduling an annual mammogram on time may experience delays in actually getting a mammogram due to back logs in mammography appointments, scheduling changes, and other unpredictable events. Because delays of 2–3 months are unlikely to alter the survival chances of women in a clinically significant way, allowing some leeway for them in compliance estimates has no obvious negative consequences and may more accurately reflect the degree to which women are attempting to comply with screening guidelines.

(b) Because these clinic and patient level delays are often beyond the control of investigators assessing the effectiveness of mammography interventions, failing to allow time for them in outcome measures may lead to false conclusions about the intervention’s effectiveness (i.e., that the intervention was not effective when in fact it ultimately was after accounting for delays). Choosing to allow 3 months rather than some other

interval for scheduling and other delays is ultimately a somewhat arbitrary decision but no more so than the decision to screen women at 12-month intervals applied in the clinical trials responsible for setting the annual screening guidelines in the first place. Because at least one previous study has already used the 15 completed month interval,3 some basis for comparison already exists.

(c) The data presented in Fig. I suggest that, at least in this population, monthly repeat mammography rates remain relatively high after 12 completed months. (If the rates dropped to zero after 12 months, the argument for a wider interval would be less defensible).

The above recommendation has implications for how reminder and recall systems are implemented for patient populations. Because lags between the clinic visit initiating a mammography referral and the actual mammography appointment can likely be estimated by clinic staff, women should be reminded to schedule their annual mammography appointments early enough so that it is possible for them to receive their mammogram on time, regardless of such delays. Because other unpredictable delays will inevitably interfere with the timely screening of even highly motivated and dedicated screeners, however, recall systems aimed at prompting women "past due" for screening should not be initiated until at least 15 completed months from the previous mammogram have elapsed.

References

Measuring mammography compliance: lessons learned from a survival analysis of screening behavior.

M R Partin, A L Casey-Paal, J S Slater, et al.