

A study on effectiveness of screening mammograms

Jian-Jian Ren^a and Petronella GM Peer^b

Background	So far, no randomized controlled trials with a mean mammographic screening interval of ≥ 2 years has demonstrated statistically significant mortality reduction for women younger than age 50. The issue of screening frequency is vital in detection of primary breast cancer.
Methods	The study group consisted of cancers diagnosed in women who participated in a serial screening programme with a mean screening interval of 2 years. To study the effectiveness of the screening, a comparison is made between the distribution of age at which the tumour could be detected when biennial mammographic screening is the only detection method, and the distribution of age at which the tumour would be detected by either biennial mammographic screening or the development of symptoms. Some recently developed statistic methods, such as bootstrap, the maximum likelihood distribution estimator for doubly censored data and the EM algorithm, are used in estimation of these distributions.
Results	The hypothesis tests and confidence intervals show that the difference between the two distributions was statistically significant for women younger than 50 and 50–70 years old, but not for women over 70 years.
Conclusions	The statistical analysis indicates that for women younger than 50, and 50–70 years of age, a screening mammogram every other year is not frequent enough to detect primary breast cancer, but for women over 70 years, it might be sufficient.
Keywords	Bootstrap, breast cancer, doubly censored data, screening mammograms
Accepted	13 March 2000

In breast cancer screening programmes, serial mammograms are available for all 'incident' cancers detected among the screenees. In addition to these screen-detected cancers, there are the so-called interval cancers, i.e. cancers diagnosed in the interval between a negative screening mammogram and the subsequent scheduled screening mammogram. Interval cancers could occur because they were missed at the previous screening examination or because tumours, initially too small to be detected by screening, had a high growth rate. Some interval cancers occurring as a result of a high growth rate would be detected at an earlier stage if screenings were done more frequently. The issue of screening frequency is especially relevant for women under 50. Studies have shown that women younger than 50 have higher growth rate cancers,⁴ thus for a screening programme to be effective it is necessary that the screenings are more frequent for women under 50 than for those over 50.⁵ Recently, two randomized controlled trials have demonstrated

statistically significant mortality reduction for women under 50.^{1–3} Both of these trials, however, used screening intervals of less than 2 years. So far no randomized controlled trials with an interval of ≥ 2 years have demonstrated statistically significant mortality reduction for women under 50.

This study analyses the effectiveness of screening mammograms with a mean screening interval of 2 years in the detection of primary breast cancer. Some recently developed statistical methods are employed to handle doubly censored observations encountered in the data set used in this research. The results of our analysis, limited to the observational data currently available, indicate that a screening mammogram every other year is not frequent enough to detect primary breast cancer in women under 50.

Data and Methods

Data

This research is based on the serial screening mammograms obtained in Nijmegen, The Netherlands, 1981–1990, which were analysed for the study of age-dependent growth rate of primary breast cancer.⁴ Data collection is described in detail elsewhere.⁴

^a Department of Mathematics, Tulane University, New Orleans, LA 70118, USA. E-mail: renj@ultral.math.tulane.edu

^b University of Nijmegen, Department of Medical Statistics, PO Box 9101, 6500 HB Nijmegen, The Netherlands. E-mail: N.Peer@MIE.KUN.NL

From the introduction of the mediolateral oblique projection in 1981 to the end of 1990, nearly 30 000 women in Nijmegen, The Netherlands, were invited for screening mammograms every other year. The percentages of these women who accepted invitations were 70% among those under 50, 60% among those aged 50–70 years, and 20% among those over 70.

For cancers diagnosed between screening examinations, i.e. the interval cancers, previous screening mammograms were re-examined by one radiologist to determine whether, in retrospect, a tumour could be identified and measured. The review showed that in several cases, a tumour nucleus shadow could be identified in the previous ‘negative’ mammogram. In other interval cancers, although there were visible tumours radiographically at the time of the diagnosis, no suspect lesions could be found in the previous ‘negative’ mammogram. An analogous review of earlier screening mammograms was also carried out for the screen-detected cancers.

In this study, the patients were eligible if at least two mammographic examinations with the mediolateral oblique projection were available. Radiographically occult cancers were excluded (16% of the interval cancers in the Nijmegen programme).⁶ They were invisible on a mammogram and showed no signs of microcalcification, thus could never be detected by mammographic screening. The resultant study group consists of 289 cancers with age at diagnosis ranging from 41 to 84 years. Among these 289 cancers, 132 were interval cancers not diagnosed at the time of screening. Of these 132 interval cancers, 79 were visible on the mammogram at the time of diagnosis, while 53 of them were palpable, but not visible on any of the mammograms (some indirect signs like microcalcification were shown on the mammogram, but no tumour shadow). The percentages of these 289 women who actually had their screening mammograms once every 2 years were 83.3% among those under 50, 84.9% among those aged 50–70, and 78.6% among those older than 70 (for interval cancers, a patient with only one previous screening examination is not considered as having had her biennial screening regularly if the time between screening and diagnosis is more than 2 years).

Statistical analysis

Consider the following two variables (in years):

A_S = age at which the tumour could be detected when biennial mammographic screening is the only detection method;

and

A_{SS} = age at which the tumour would be detected by either biennial mammographic screening or the development of symptoms,

where the index ‘S’ is for ‘screening’ and ‘SS’ for ‘screening or symptoms’. One should note that A_S only includes the screen-detected cancers, while A_{SS} includes screen-detected cancers and interval cancers. Let F_S and F_{SS} be the distribution functions of A_S and A_{SS} , respectively. A comparison between F_S and F_{SS} may be conducted to study the effectiveness of the screening mammograms in the detection of primary breast cancer. If m_S and m_{SS} are the medians of F_S and F_{SS} , respectively, then a big difference between m_S and m_{SS} implies that the age at which the tumour could be detected using biennial mammographic screening as the only detection method is much later than the age at which the tumour developed. One may note that $m_S > m_{SS}$

should always hold based on the above definitions of A_S and A_{SS} , and relevant inference in this study is the size of the difference between m_S and m_{SS} . A statistically significant difference between m_S and m_{SS} , say, $(m_S - m_{SS}) > \Delta$ for some $\Delta > 0$, would indicate that the screening was not done frequently enough to detect the cancer.

Among 289 cancers in the study group, 45 women had tumour volumes observed at their first screening mammograms. Some of these women, though not treated as interval cancers, had positive mammograms at their first screenings because of the retrospective examination due to interval cancers, while the rest of them continued their biennial mammograms after their positive first screenings. If these women had started their screening earlier and continued having mammograms every 2 years, the tumour could have been (at least very likely) observed before their first mammograms. Thus we assume that A_S for these 45 individuals was less than the age at the first screening mammogram, and so they yield 45 left censored observations. A *left censored observation* is an available quantity which is larger than the desired, but unavailable, quantity (A_S in the case considered here). Similarly, 132 of the 289 cancers did not have tumour volume observed at the last available screening mammogram, yielding 132 right censored observations (for each individual, A_S was greater than the age at the last screening examination). During the serial screening mammograms, 112 were observed tumour growth yielding uncensored observations (for each individual, A_S was actually observed). All these observations together consist of a *doubly censored data set*, which is used to estimate the distribution of A_S . Intuitively, the use of the left censored or right censored observation in this problem takes into account that, for some individuals, A_S is, with positive probability, less than or greater than the observation we have. Thus, it gives a more accurate estimation for F_S . The estimator for F_S , denoted as \hat{F}_S , is called the *non-parametric maximum likelihood estimator* and its computation and asymptotic properties are studied in recent statistics literature.⁷

One may notice that for those 132 right censored observations in the doubly censored model of A_S , we assume that A_S could have been observed sometime after the last screening if the screening had been conducted continuously every 2 years. These 132 cases are all these interval cancers. However, this does not mean that we assume A_S is definitely observable at the next regular screening. The reasons for this are that not all observed interval cancers occurred before the next scheduled screening (some patients did not have their screenings on schedule), and that, as mentioned earlier, among all interval cancers, 53 did not show any tumour nucleus shadow on the mammograms at the time of diagnosis. These might have not been detected by the screening even if the next screening mammograms were conducted on time in these cases. Thus it is not appropriate to assume that all interval cancers can be observed at the next scheduled screening.

Another point worth mentioning is that since the screening interval was fixed as 2 years, there is no information available on the exact time at which the tumour developed between a negative screening mammogram and the subsequent positive mammogram. The definition of A_S is not equivalent to the time at which the tumour developed, thus A_S is not subject to any censoring between a negative screening mammogram and the

Table 1 Tests and confidence intervals on $(m_S - m_{SS})$

Age group	Total no. of observations	Median of \hat{F}_S	Median of \hat{F}_{SS}	90% CI for $m_S - m_{SS}$	Δ in Test (1) (in months)	T
Age <50	48	48.07	45.50	(0.68, 4.46)	7	1.73
					8	1.65
					9	1.58
Age 50–70	185	63.59	59.80	(2.26, 5.31)	26	1.74
					27	1.66
					28	1.57
Age >70	56	74.96	74.13	(-1.33, 2.99)	0	0.63

subsequent positive mammogram, and there were in total 112 such cases in our study for A_S .

To estimate the distribution of A_{SS} , the 45 left censored observations above are still used as left censored observations, but the rest of 244 cancers, yielding uncensored observations (for each individual, A_{SS} was actually observed), were either detected at screening or clinically diagnosed based on symptoms. These observations together consist of a left censored data set, and are used to compute the reversed Kaplan-Meier estimator⁷ for F_{SS} , denoted as \hat{F}_{SS} .

To compare the distributions F_S and F_{SS} , the following hypothesis test is conducted for women under 50, women aged 50–70, and women over 70, respectively:

$$H_0: (m_S - m_{SS}) \leq \Delta \text{ versus } H_1: (m_S - m_{SS}) > \Delta. \quad (1)$$

Our goal is seeking sufficient evidence to conclude the alternative hypothesis H_1 with some $\Delta > 0$ for different age groups. For this test, the test statistic is computed by

$$T = \frac{\sqrt{n} \{[(\text{Median of } \hat{F}_S) - (\text{Median of } \hat{F}_{SS})] - \Delta\}}{SE}$$

where n is the total number of observations used to compute \hat{F}_S and \hat{F}_{SS} , and SE is the standard error of

$$\sqrt{n} \{[(\text{Median of } \hat{F}_S) - (\text{Median of } \hat{F}_{SS})] - \Delta\}.$$

We know⁹ that for large n , T is asymptotically standard normal, SE can be estimated by the bootstrap method^{10,11} (1000 bootstrap samples are used in this study), and the rejection region for 95% significance level is $T > 1.645$.

It should be pointed out that the complexity of this current analysis is due to the doubly censored data naturally encountered for the estimation of the distribution of A_S . In recent years, new statistical methods^{8–9} have been developed precisely for this type of data.

Results

In Table 1, it is easy to see that differences between the medians for those under 50 and those aged 50–70 are 2.57 and 3.79 years, respectively, which is greater than the mean screening interval of 2 years. This could occur because for those 132 right censored observations in the doubly censored model of A_S , it is assumed that A_S could have been observed sometime after the last screening if the screening had been conducted continuously every 2 years, but not necessarily at the next scheduled screening. Thus, large differences in the medians between A_S and A_{SS} correspond to frequent occurrences of interval cancers in these age groups.

Discussion and Conclusions

In Table 1, it is clear that with a 95% significance level, the largest values of Δ which lead to rejection of the null hypothesis H_0 in favour of H_1 in (1) for women under 50 and those aged 50–70 years are 8 months and 27 months, respectively. The positive left endpoints of 90% confidence intervals for the difference between m_S and m_{SS} in these two age groups are consistent with the test results. This means that there are statistically significant differences between m_S and m_{SS} in these two age groups, which indicate that for these two age groups, a biennial screening mammogram is not frequent enough to detect primary breast cancer effectively. On the other hand, for $\Delta = 0$ in (1), the test statistic for those over 70 is 0.63 and the left endpoint of the 90% confidence interval of $(m_S - m_{SS})$ is negative indicating that for women over 70 years, a screening mammogram every other year might be sufficient.

As summarized earlier, some recent studies^{1–3,10} have demonstrated a statistically significant mortality reduction for women under 50, but so far no trials with an interval of ≥ 2 years have demonstrated the same fact. This current analysis indicates that biennial mammographic screening is not frequent enough to detect primary breast cancer in women under 50.

However, it must be noted that due to the limits of the observational data we have, cost-benefit and other risk factors (e.g. radiation exposure) are not taken into account in our analysis, and further study should be conducted to determine how frequently younger women should have their screening mammograms in order to detect primary breast cancer effectively.

Acknowledgement

Research partially supported by NSF Grants DMS-9510376 and DMS-9626532/9796229.

References

- Bjurstaam N, Björnelid L, Duffy S *et al.* The Gothenburg breast screening trial. *Cancer* 1997;**80**:2091–99.
- Anderson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmö mammographic screening program. *Monogr Natl Cancer Inst* 1997;**22**:63–67.
- Hendrick RE, Smith RA, Rutledge JH III, Smart CR. Benefit of screening mammography in women aged 40–49: a new meta-analysis of randomized controlled trials. *Monogr Natl Cancer Inst* 1997;**22**:87–92.
- Peer PG, Van Dijck JA, Hendriks JH, Holland R, Verbeek ALM. Age-dependent growth rate of primary breast cancer. *Cancer* 1993;**71**:3547–51.

- ⁵ Wolfe JN. Breast parenchymal patterns and their changes with age. *Radiology* 1976;**121**:545–52.
- ⁶ Peeters PHM, Verbeek ALM, Hendriks JHCL, Holland R, Mravunac M, Vooijs GP. The occurrence of interval cancers in the Nijmegen screening programme *Br J Cancer* 1989;**59**:929–32.
- ⁷ Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Amer Statist Assoc* 1958;**3**:457–481.
- ⁸ Mykland P, Ren J. Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data. *Ann Stist* 1996;**24**:1740–64.
- ⁹ Ren J, Zhou M. L-estimators and M-estimators with doubly censored data. *J Nonparametric Stist* 1997;**8**:1–20.
- ¹⁰ Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.
- ¹¹ Bickel PJ, Ren J. The m out of n bootstrap and goodness of fit tests with doubly censored data. *Robust Statistics, Data Analysis and Computer Intensive Methods*, Lecture Notes in Statistics, Springer-Verlag, 1996; **109**:35–47.
- ¹² Feig SA. Increased benefit from shorter screening mammography intervals for women ages 40–49 years. *Cancer* 1997;**80**:2035–39.