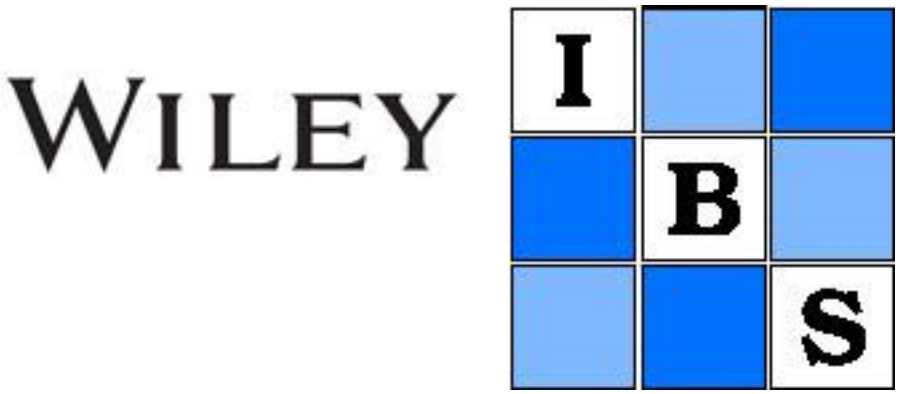

BIOMETRICS 54, 1569–1578
December 1998

Evaluating the Age to Begin Periodic Breast Cancer Screening Using Data from a Few Regularly Scheduled Screenings

Stuart G. Baker

Biometry Branch, Division of Cancer Prevention, National Cancer Institute, EPN 344, 9000 Rockville Pike, Bethesda, Maryland 20892, U.S.A.

SUMMARY

To evaluate various ages to begin periodic breast cancer screening, we propose a method of analysis that can be applied to either a nonrandomized or a randomized study involving only a few screenings at regular intervals. For the analysis of data from a nonrandomized study, we assume (i) once breast cancer can be detected on screening and confirmed by biopsy, it will stay that way; (ii) given age, the probability of breast cancer detection does not depend on year of birth; and (iii) subjects who refuse screening have the same rates of breast cancer mortality following diagnosis as screened subjects had they not received screening. The key idea is that older screened subjects are controls for younger screened subjects. For the analysis of data from a randomized study, we relax assumption (iii). Based on the HIP randomized trial and assumptions (i) and (ii), we estimate that starting periodic breast cancer screening with mammography and physical examination at age 40 instead of age 50 reduces breast cancer mortality by 14 per 10,000 with a 95% confidence interval of (−4/10,000, 32/10,000). This must be weighed against an estimated increase in the number of biopsies that do not detect cancer of 580 per 10,000 with a 95% confidence interval of (520/10,000, 650/10,000).

1. Introduction

An important public health question is what the reduction in breast cancer mortality is associated with various ages to begin periodic screening for breast cancer, with most of the controversy centering around age 40 versus age 50 (e.g., Fletcher et al., 1993). Although the most definitive answer would come from trials that randomize subjects to start periodic screening at various ages, such trials are very expensive and none have been completed. Meta-analyses of completed randomized trials are difficult to interpret because the trials compare an offer of screening with no screening rather than different ages to start periodic screening (e.g., Kerlikowske et al., 1995). We develop an alternative approach to address this question by analyzing data from either a nonrandomized or randomized study involving a few screenings. The methodology extends and simplifies periodic screening evaluation (Baker, 1989; Baker and Chu, 1990) by obviating the original EM algorithm and bootstrap, smoothing over ages to tighten confidence intervals, and providing new insight into the key idea of using older subjects on the first screening as controls for younger subjects on the first screening.

We illustrate the methodology using data from the HIP (Health Insurance Plan of New York) randomized trial (Shapiro et al., 1988) of breast cancer screening with physical examination and mammography. Approximately 60,000 women were randomized to either a control group not offered screening or a study group offered four annual screenings. Among the study group, approximately two thirds accepted at least the initial screening and one third refused; we call them acceptors and refusers, respectively. We analyze the data three ways: a nonrandomized study based only on data

Email address: sb16i@nih.gov

Key words: Noncompliance; Nonrandomized trials; Observational studies; Periodic screening evaluation; Poisson regression.

from acceptors and refusers, a randomized study relaxing one assumption, and a randomized study relaxing two assumptions.

The paper is organized as follows. Section 2 describes estimation from a nonrandomized study, and Section 3 describes estimation from a randomized trial. Section 4 shows how to combine parameter estimates to estimate the reduction in breast cancer mortality due to starting periodic screening at a younger age; it also discusses estimation of the increase in biopsies that do not detect cancer. Section 5 briefly discusses the use of a short-term endpoint, and Section 6 presents results from several analyses of the HIP data.

2. Estimates from a Nonrandomized Study of a Few Screenings

Subjects who are asymptomatic with no prior screening are offered at least two screenings at intervals of constant length, say 1 year. Their age range should be the same as the range of ages for evaluating the start of periodic screening. For example, to compare periodic screening starting at age 40 versus 60, we require subjects ages 40 to 60 to receive screening. Data are collected from both refusers and acceptors. Follow-up to determine breast cancer mortality occurs only in subjects detected with breast cancer up to and including the time of the last screening. We require the definition and assumptions that follow.

DEFINITION: *A screening in the HIP study is positive if either mammography or physical examination is positive and it is confirmed by biopsy.*

ASSUMPTION 1: *Once a screening can detect breast cancer that is confirmed by biopsy, it will always detect breast cancer that is confirmed by biopsy.*

In other words, there is a continuous period of time, up to detection resulting from symptoms, when breast cancer can be detected on screening and confirmed by biopsy. (Prior to this period of time, there may be a false negative test defined as when mammography and physical examination are negative, but a biopsy, if done, would have been positive. Also prior to this period of time, there may be a false positive test, also called an unnecessary biopsy, when mammography or physical examination are positive and the biopsy is negative.) Assumption 1 most likely holds when nonmicroscopic tumors are required for detection on biopsy, as in the HIP study.

ASSUMPTION 2: *Given age, the probability of breast cancer detection does not depend on year of birth.*

ASSUMPTION 3: *Breast cancer mortality rates following detection of breast cancer due to symptoms are the same in refusers as in subjects who received screening had they not been screened.*

2.1 *Age-Specific Probabilities of Cancer Detection*

Let a denote the age at the start of the study. In the simplest design, asymptomatic subjects are scheduled for screenings at ages a and $a+1$ (Figure 1, left) and breast cancer can be detected in the following ways:

$F =$ positive first screening at age a,
$I =$ negative screening at age a and symptomatic cancer in the following 1-year interval prior to the screening at age $a+1$,
$S =$ negative screening at age a, no symptomatic cancer in the following 1-year interval, and positive on the subsequent screening at age $a+1$.

Because F, I, and S are observed in study subjects, we can directly estimate their age-specific probabilities. To evaluate periodic cancer screening, we also need to estimate the age-specific probability of symptomatic detection if the same subjects had not been screened. To this end, we consider a thought experiment in which the same asymptomatic subjects age a are not screened at age a but are scheduled for a first screening at age $a+1$ (Figure 1, right), so breast cancer can be detected in the following ways:

$N =$ following no screening at age a, symptomatic cancer in the 1-year interval prior to the screening at age $a+1$,
$L =$ following no screening at age a, no symptomatic cancer in the 1-year interval, and a positive first screening at age $a+1$.

Let $q(d \mid a)$ denote the probability of type $d \in \{F, I, S, N, L\}$ detection at age a. As derived in Appendix I, under Assumption 1, the age-specific probability of breast cancer detection if screened subjects had not received screening is

$$q(N \mid a) = q(F \mid a) + q(S \mid a) - q(F \mid a+1), \tag{1}$$

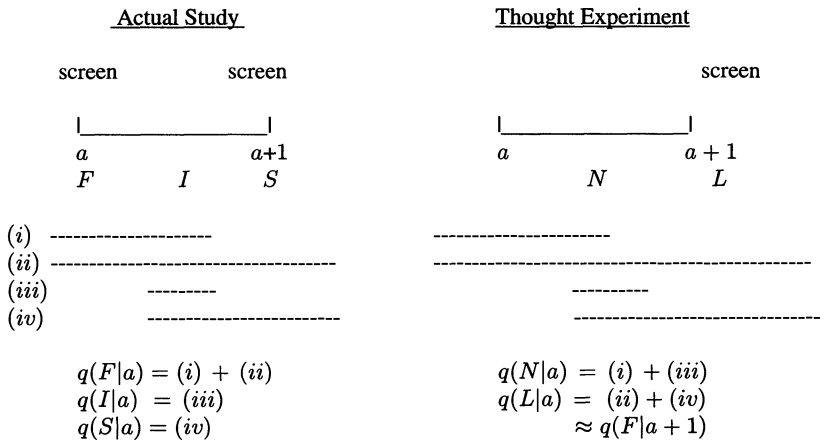

Figure 1. Schematic derivation of age-specific probability of type N detection. The dashed line indicates the period when a screening would detect cancer confirmed by biopsy. Under Assumption 1, this period is continuous and ends with symptomatic detection. The lines (i), (ii), (iii), (iv) define whether or not the period began before age a and ended before age $a + 1$.

where $q(F \mid a + 1)$ is the probability of a positive first screening 1 year after the start of the study among subjects age a at the start of the study who would be age $a + 1$ at the first screening. Because $q(F \mid a + 1)$ applies only to the thought experiment, it cannot be estimated from data in the actual study in which subjects age a at the start of the study receive a first screening at age a. To estimate $q(F \mid a + 1)$ from data in the actual study, we invoke Assumption 2 and use data from subjects age $a + 1$ at the start of the study who receive a first screening at age $a + 1$. This is what we mean by using older subjects as controls for younger subjects.

Due to sampling variability, some estimates of $q(N \mid a)$ may be negative. To reduce sampling variability, we use data from multiple screenings to estimate $q(I \mid a)$ and $q(S \mid a)$. (See Appendix I.) We also smooth estimates over age in the following manner. Let k_{aj} denote the number of asymptomatic subjects age a at screening j. Let n_{daj} denote the number of subjects with type d detections given they are asymptomatic at age a and screening j. (The data are from Table 1 in Baker and Chu (1990).) Unlike Baker and Chu (1990), we cannot use data from subjects positive after skipping a screening, but this should not substantially bias estimates as there are only three such subjects and the data are not directly informative concerning the age-specific probabilities of types F, I, S, or N detection. Because breast cancer is rare, we assume the number of F, I, and S detections follow a Poisson distribution, $n_{Fa0} \sim \text{Poisson}(k_{a0}q(F \mid a))$, and $n_{Iaj} \sim \text{Poisson}(k_{aj}q(I \mid a + j))$, and $n_{Saj} \sim \text{Poisson}(k_{aj}q(S \mid a + j))$. To smooth estimates over age, we introduce the parameterization $q(F \mid a) = \exp(\beta_{F0} + \beta_{F1}a + \beta_{F2}a^2), q(I \mid a + j) = \exp(\beta_{I0} + \beta_{I1}(a + j) + \beta_{I2}(a + j)^2), q(S \mid a + j) = \exp(\beta_{S0} + \beta_{s1}(a + j) + \beta_{S2}(a + j)^2)$. There was no evidence to reject $\beta_{F2} = 0, \beta_{I2} = 0$, or $\beta_{IS} = 0$, with differences in deviances of .8, 3.0, and .4 respectively. The age-specific probability of F was largest, followed by S, then I (Figure 2, left). Substituting the regression fits for F, S, and I into (1) gives a smoothed estimate of the age-specific probability of N (Figure 2, top right).

2.3 *Breast Cancer Mortality Following Detection*

For evaluation, we estimate the probability of breast cancer death following type d detection at age a, which we denote $m(d, a)$. Using a follow-up period of 15 years, we compute $m(d, a) = \Sigma_{u=1}^{15} surv(a, u) \Pi_{j=1}^{u-1} (1 - \hat{\lambda}_{dj})\hat{\lambda}_{du}$, where $surv(a, u)$ is the known probability of surviving competing risk from age a to $a + u$ and $\hat{\lambda}_{du}$ is the cancer mortality rate at year u after type d detection. We estimate $\hat{\lambda}_{du}$ by the number of deaths in interval u since detection divided by the number at risk at the start of interval u. For $d = F, I$, and S, data are obtained directly. For $d = N$, we invoke Assumption 3 and use data from refusers. Because I and N occur between screenings at ages a and $a + 1$, we average the probabilities for ages a and $a + 1$.

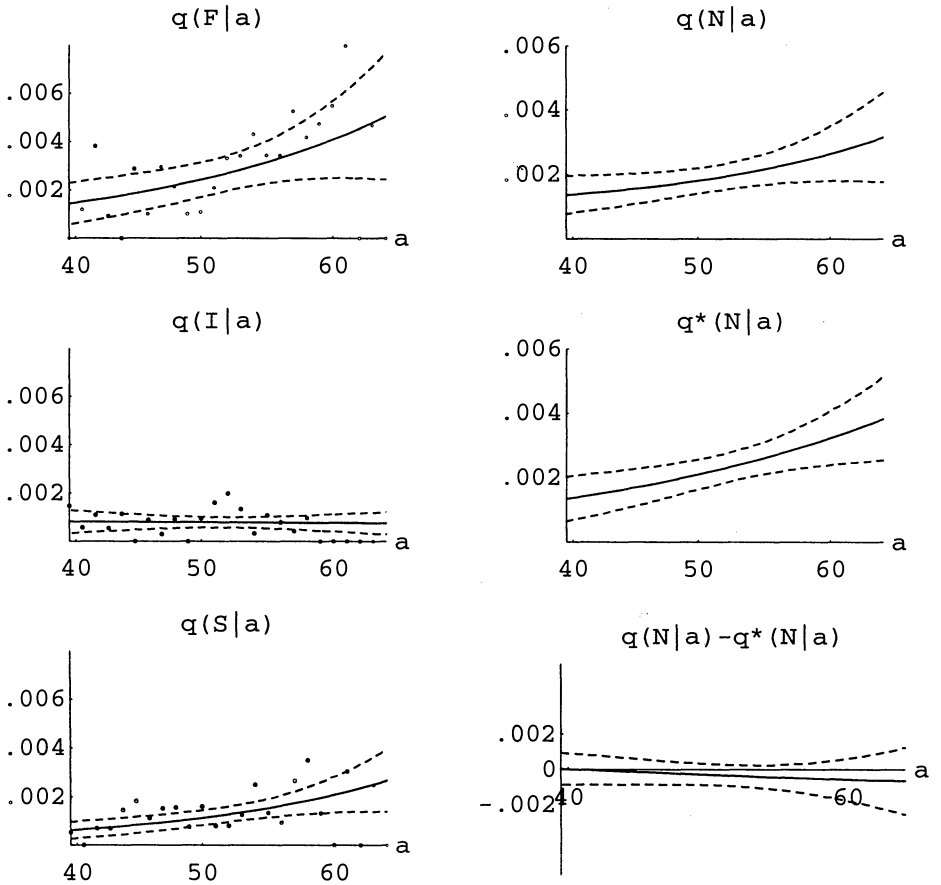

Figure 2. Regression estimates of the age-specific probabilities of various types of detection. The points represent the proportion detected at each age, the solid lines represent the estimated curves from the Poisson regression, and the dashed lines represent the upper and lower bounds of the 95% confidence intervals. All estimates are from the nonrandomized analysis except that for $q^*(N \mid a)$.

3. Estimates from a Randomized Trial of a Few Screenings

In this framework, the major advantage of a randomized trial is the elimination of Assumption 3. Let γ denote the fraction of subjects with cancer who accepted screening, γ_{Ct} the breast cancer mortality rate in randomized controls, λ_{Rt} the breast cancer mortality rate in refusers, and λ^*_{Nt} the breast cancer mortality rate if subjects screened had not received screening. For a few years after randomization, $\hat{\lambda}_{Ct} \approx \hat{\lambda}^*_{Nt}\gamma + \hat{\lambda}_{Rt}(1-\gamma)$. (See Baker (1997) for a review of this type of estimate.) Therefore, the mortality rate among acceptors at time t if they had not been screened is $\hat{\lambda}^*_{Nt} \approx [\hat{\lambda}_{Ct} - (1-\gamma)\hat{\lambda}_{Rt}]/\gamma$.

Similarly, let π denote the fraction of subjects who accepted screening. By a similar argument, we can estimate $q(N \mid a)$ as $\hat{q}^*(N \mid a) = [\hat{q}(C \mid a) - (1-\pi)\hat{q}(R \mid a)]/\pi$ and eliminate Assumption 1. We define a type I randomized analysis as an analysis using $\hat{\lambda}^*_{Nt}$ instead of $\hat{\lambda}_{Nt}$; we define a type II randomized analysis as an analysis that also uses $\hat{q}^*(N \mid a)$ instead of $\hat{q}(N \mid a)$.

4. Evaluating the Age to Begin Periodic Screening

4.1 *Reduction in Breast Cancer Mortality*

By combining parameter estimates, we estimate the reduction in cancer mortality due to starting periodic cancer screening at a younger age. For asymptomatic subjects age a, the probability of breast cancer death from waiting until age $a+k$ and then starting periodic screening is $p_{a+k} + \phi$, where $p_{a+k} = \Sigma_{j=0}^{k}\, \hat{q}(N \mid a+j)m(N, a+j) + \hat{q}(F \mid a+k+1)m(F, a+k+1)$ and ϕ is the probability of breast cancer mortality from breast cancers detected after a negative screening at age $a+k$.

Similarly, for the same asymptomatic subjects age a, the probability of breast cancer death from starting periodic screening at age a is $p_a + \phi$, where $p_a = \hat{q}(F \mid a)m(F, z) + \Sigma_{j=0}^{k} \hat{q}(I \mid a+j)m(I, a+j) + \Sigma_{j=0}^{k} \hat{q}(S \mid a+j)m(S, a+j)$. The reason ϕ is the same in both equations is that it applies to the same subset of subjects with breast cancer after a negative screening at age $a+k$ because, under Assumption 1, they would have been, or were negative, on all previous screenings. The estimated reduction in mortality is $g = p_a - p_{a+k}$, with the asymptomatic variance calculated using the delta method (Appendix II).

4.2 *Increase in Unnecessary Biopsies*

Any reduction in cancer mortality associated with periodic cancer screening must be weighed against an increase in unnecessary biopsies, i.e., biopsies that do not detect cancer. Let ρ_F and ρ_S denote the probability of an unnecessary biopsy on the initial screening and subsequent screenings, respectively. The expected number of unnecessary biopsies over k screenings is $\rho_F + (k-1)\rho_s$ with asymptotic variance $\rho_F(1-\rho_F)/w_F + (k-1)^2\rho_S(1-\rho_S)/w_S$, where w_F and w_S are the number of initial and subsequent screenings, respectively. Based on HIP data (Strax et al., 1970, Tables 1 and 5), $\hat{\rho}_F = .0125$, $w_F = 20{,}211$, $\hat{\rho}_S = .0051$, and $w_S = 44{,}056$, so the estimated number of unnecessary biopsies per 10,000 for 10 yearly screenings is 580 with 95% confidence interval (520, 650). Based on more recent data (Sickles, 1997), $\hat{\rho}_F = .014$, $w_F = 29{,}694$, $\hat{\rho}_S = .006$, and $w_S = 42{,}451$, which gives an estimate per 10,000 of 680 with 95% confidence interval (610, 750).

5. Extension to a Short-Term Endpoint

When screening technology is changing rapidly over time, investigators may wish to investigate the effect of screening on a short-term endpoint, such as the probability of a positive node, i.e., the probability that cancer has spread to a lymph node at the time of detection. Although a reduction in the probability of a positive node does not necessarily imply a reduction in breast cancer mortality, it is suggestive, and similarly for no reduction. By modifying Assumption 3 and the appropriate formulas, the methodology is readily adapted.

6. Results

To compare periodic breast cancer screening starting at age 40 versus age 50 and age 50 versus age 60, we analyzed data from the HIP study three ways: a nonrandomized study, which ignores data from controls; a type I randomized analysis, which relaxes Assumption 3; and a type II randomized analysis, which relaxes both Assumptions 1 and 3 (Table 1). The nonrandomized and type I randomized analyses gave similar estimates, supporting Assumption 3. The type II randomized analysis gave larger estimates, which may make Assumption 1 questionable. However, the evidence to reject Assumption 1 is weak. A plot of the difference between $\hat{q}^*(N \mid a)$ and $\hat{q}(N \mid a)$ included zero in the 95% confidence interval for most ages (Figure 2, right bottom). The fact that the difference decreased with age without turning upward is related to the use of a linear model in the exponent for age. If Assumption 1 were violated, one would expect that, due to some positive cancers becoming negative, $\hat{q}^*(N \mid a)$ would be smaller than $\hat{q}(N \mid a)$, which was not the case. Although the type II analysis may reduce bias, the variance increases substantially because subjects were not their own controls.

Table 1

Estimated reduction in probability of positive lymph nodes and cancer mortality due to starting periodic cancer screening at various ages

Age to start of periodic screening	Analysis	Decrease in positive lymph nodes per 10,000[a]	Decrease in breast cancer deaths per 10,000[a]
40 vs. 50	Nonrandomized	23 (11, 36)	11 (−7, 29)
	Randomized, type I	25 (14, 37)	14 (−4, 32)
	Randomized, type II	31 (−3, 64)	20 (−20, 59)
50 vs. 60	Nonrandomized	39 (24, 54)	25 (1, 49)
	Randomized, type I	41 (28, 55)	28 (4, 51)
	Randomized, type II	62 (33, 92)	49 (11, 86)

[a] 95% confidence interval in parentheses.

For a one-sided test of the null hypothesis of no effect of starting periodic screening at age 50 instead of 40, the p-value was .11, .06, and .16 for the nonrandomized, type I randomized, and type II randomized analyses, respectfully. Thus, our analyses suggest a possible reduction in mortality due to starting periodic screening at 40 instead of 50, which is approximately half that from starting periodic screening at 50 instead of 60. Our results are not directly comparable with those of other analysts of the HIP data because we are trying to answer a different question. For example, Chu, Smart, and Tarone (1988) computed a p-value of .02 for a one-sided test of no effect of an offer of four screenings between ages 40 and 50. In any case, whether or not periodic breast cancer screening is deemed worthwhile also depends on the trade-off between reduction in mortality and unnecessary biopsies, the attitude toward uncertainty, results of other studies, and changes in screening technology since the HIP study.

7. Discussion

Most models to evaluate cancer screening are based on parameters describing the natural history of cancer and the characteristics of the screening tests (e.g., Brookmeyer, Day, and Moss, 1986; Flehinger and Kimmel, 1987; Habbema et al., 1984; Stevenson, 1995; van Oortmarssen, Boer, and Habbema, 1995; Day and Walter, 1984; Louis, Albert, and Heghnian, 1978; Shwartz, 1978; Stevenson, 1995; Walter and Day, 1983; Zelen, 1993; Zelen and Feinleib, 1969). These models usually require strong assumptions about underlying parameters involving the incidence of preclinical cancer, the duration of the preclinical cancer, the sensitivity of the screening test, and, especially, the effect of early intervention on cancer mortality. In contrast, our approach requires different assumptions, which may be more plausible. In the spirit of Baker and Chu (1990) and Flanders and Longini (1990), many underlying parameters are not specified in the model, although they have an indirect effect on estimation. For example, the probability that mammography or physical examination is positive in a subject with cancer is not in the model but affects the probability of a positive screening.

We believe that, if the assumptions hold, our nonrandomized design has advantages over other methods for the analysis of observational screening data. By assuming negative screenings cannot follow positive screenings and that there is no effect of birth cohort given age, we use older screened subjects as controls for younger screened subjects. This eliminates selection bias in estimating cancer incidence had subjects who were screened not received screening. This bias arises in case–control studies (e.g., Weiss, McKnight, and Stevens, 1992) and the use of population controls (Morrison, Brisson, and Khalid, 1988). There is still the possibility of selection bias in estimating breast cancer mortality rates following detection had subjects who were screened not received screening. However, at least in the HIP analysis, this bias appeared to be small.

Some assumptions can be eliminated if investigators are willing to randomize subjects. Unlike other randomized trials, the trial can involve as few as two screenings. If investigators are interested in the probability of node positive, the trial can be completed within a year. Although we estimated the probability of unnecessary biopsies, we could also extend the methodology to estimate the monetary costs of screening (e.g., Mushlin and Fintor, 1992). We hope to apply the methodology to data sets for breast cancer screening involving only mammography or only breast self-examination. Although the methodology would not be appropriate for cervical cancer screening because negative screenings may follow positive screenings, it may be applicable to lung cancer screening.

Calculations can be reproduced using software written in Mathematica 3.0, which is available at http://www.dcp.nci.nih.gov/BB/software.html.

Acknowledgements

This paper has benefited from comments following presentations and on earlier versions. The author thanks William Black, Kenneth Chu, Robert Connor, Sylvan Green, Ruth Etzioni, Larry Freedman, Franka Loeve, Mark Moran, Philip Prorok, Hans Trampisch, and Martin Weinrich.

Résumé

Pour évaluer les âges auxquels il faut commencer le dépistage périodique du cancer du sein, on propose une méthode d'analyse pouvant être appliquée dans des études portant aussi bien sur des échantillons aléatoires que non aléatoires dans le cadre de stratégies de dépistage impliquant peu d'examens de dépistage à intervalles réguliers. Pour l'analyse des données à partir d'échantillon non aléatoires on suppose 1) dès que le cancer du sein est détecté à l'examen de dépistage et qu'il est confirmé par une biopsie, le résultat est définitivement acquis, 2) pour un âge particulier la probabilité de détecter un cancer du sein ne dépend pas de l'année de naissance, 3) les taux de mortalité par cancer du sein chez les sujets après un diagnostic symptomatique sont les mêmes

chez les sujets qui refusent le dépistage que chez les sujets ayant reçu le dépistage et qui n'ont pas été détectés. L'idée de base consiste à prendre les sujets ayant eu un dépistage à un âge élevé comme groupe de contrôle pour les sujets ayant eu un dépistage à un âge jeune. Pour l'analyse des données portant sur un échantillon aléatoire, l'hypothèse 3 est abandonnée. En se fondant sur les essais randomisés du HIP (Health Insurance Plan of New York) et sur les deux premières hypothèses, on estime à quel âge, 40 ans ou 57 ans, il faut commencer le dépistage périodique avec une mammographie et un examen physique pour réduire le taux de mortalité par cancer du sein de 13/10000 avec un intervalle de confiance de [1/10000, 28/10000]. Ceci doit être discuté en considérant parallèlement l'accroissement estimé du nombre de biopsies négatives: 580 cas/10000, avec un intervalle de confiance allant de 520 cas à 650 cas pour 10000.

References

Baker, S. G. (1989). Innovations in screening: Evaluating periodic screening without using data from a control group. In *Advances in Cancer Control,* Volume VI, P. F. Engstrom, P. Anderson, and L. Mortenson (eds), 15–21. New York: Alan R. Liss.

Baker, S. G. (1997). Compliance, all-or-none. In *The Encyclopedia of Statistical Science,* Update Volume 1, S. Kotz, C. R. Read, and D. L. Banks (eds), 134–138. New York: Wiley and Sons.

Baker, S. G. and Chu, K. C. (1990). Evaluating screening for the early detection and treatment of cancer without using a randomized control group. *Journal of the American Statistical Association* **85,** 321–327.

Brookmeyer, R., Day, N. E., and Moss, S. (1985). Case-control studies for estimation of the natural history of preclinical disease from screening data. *Statistics in Medicine* **5,** 127–138.

Chu, K. C., Smart, C. R., and Tarone, R. E. (1988). Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan Clinical Trial. *Journal of the National Cancer Institute* **80,** 1125–1132.

Day, N. E. and Walter, S. D. (1984). Simplified models for screening for chronic disease: Estimation procedures for mass screening programmes. *Biometrics* **40,** 1–14.

Flanders, W. D. and Longini, I. M. (1990). Estimating benefits of screening from observational cohort studies. *Statistics in Medicine* **9,** 969–980.

Flehinger, B. J. and Kimmel, M. (1987). The natural history of lung cancer in a periodically screened population. *Biometrics* **43,** 127–144.

Fletcher, S. W., Black, W., Harris, R., Rimer, B. K., and Shapiro, S. (1993). Report of the International Workshop on Screening for Breast Cancer. *Journal of the National Cancer Institute* **85,** 1644–1656.

Habbema, J. D. F., van Oortmarssen, G. J., Lubbe, J. Th. N., and van der Maas, P. J. (1984). The MISCAN simulation for the evaluation of screening for disease. *Computer Methods and Programs in Biomedicine* **20,** 79–83.

Kerlikowske, K., Grady, D., Rubin, S. M., Sandrock, C., and Ernster, V. L. (1995). Efficacy of Screening Mammography: A Meta-analysis. *Journal of the American Medical Association* **273,** 149–154.

Louis, T. A., Albert, A., and Heghnian, S. (1978). Screening for the early detection of cancer. III. Estimation of disease natural history. *Mathematical Biosciences* **40,** 111–144.

Morrison, A. S., Brisson, J., and Khalid, H. (1988). Breast cancer incidence and mortality in the breast cancer detection demonstration project. *Journal of the National Cancer Institute* **80,** 1540–1547.

Mushlin, A. I. and Fintor, M. A. (1992). Is screening for breast cancer cost-effective? *Cancer* **69,** 1957–1962.

Shwartz, M. (1978). An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease. *Cancer* **41,** 1550–1564.

Shapiro, S., Venet, W., Strax, P., and Venet, L. (1988). *Periodic Screening for Breast Cancer, the Health Insurance Plan Project and Its Sequelae.* Baltimore: Johns Hopkins University Press.

Sickles, E. A. (1997). Breast cancer screening outcomes in women age 40–49: Clinical experience with service screening using mammography. *Monographs of the National Cancer Institute* **22,** 99–104.

Stevenson, C. E. (1995). Statistical models for cancer screening. *Statistical Methods in Medical Research* **4,** 18–32.

Strax, P., Venet, L., Shapiro, S., Gross, S., and Venet, W. (1970). Breast cancer found on repetitive examination in mass screening. *Archives of Environmental Health* **20,** 761.

Van Oortmarssen, G. J., Boer, R., and Habbema, J. D. F. (1995). Modelling issues in cancer screening. *Statistical Methods in Medical Research* **4,** 33–54.

Walter, S. D. and Day, N. E. (1983). Estimation of the duration of a pre-clinical disease state using screening data. *American Journal of Epidemiology* **118,** 865–886.

Weiss, N., McKnight, B., and Stevens, N. (1992). Approaches to the analysis of case–control studies of the efficacy of screening for cancer. *American Journal of Epidemiology* **135,** 817–823.

Zelen, M. (1993). Optimal scheduling of examination for the early detection of disease. *Biometrika* **80,** 279–293.

Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic disease. *Biometrika* **56,** 601–614.

Received December 1995; *revised September* 1997; *accepted October* 1997.

Appendix I

Age-Specific Incidence of Type N Detection

Let $X_a = 1$ if a screening at age a would be positive and 0 otherwise, and let $Y_a = 1$ if symptomatic cancer would be detected by age a in the absence of screening and 0 otherwise. Under the model, the age-specific probabilities of detection are $q(F \mid a) = \text{pr}(X_a = 1 \mid Y_a = 0)$, $q(I \mid a) = \text{pr}(Y_{a+1} = 1, X_a = 0 \mid Y_a = 0)$, $q(S \mid a) = \text{pr}(X_{a+1} = 1, Y_{a+1} = 0, X_a = 0 \mid Y_a = 0)$, $q(N \mid a) = \text{pr}(Y_{a+1} = 1 \mid Y_a = 0)$, and $q(L \mid a) = \text{pr}(X_{a+1} = 1, Y_{a+1} = 0 \mid Y_a = 0)$. As a first step, we write the age-specific probability of type N detection as

$$\begin{aligned} q(N \mid a) &= \text{pr}(Y_{a+1} = 1, X_a = 0 \mid Y_a = 0) + \text{pr}(Y_{a+1} = 1, X_a = 1 \mid Y_a = 0) \\ &= q(I \mid a) = q(F \mid a) - q(F0 \mid a), \end{aligned} \tag{A.1}$$

where $q(F0 \mid a) = \text{pr}(Y_{a+1} = 0, X_a = 1 \mid Y_a = 0)$ is the unobserved probability of no cancer in the first interval in the absence of screening but positive if a first screening were at age a. To eliminate $q(F0 \mid a)$, we write the age-specific probability of type L detection as

$$\begin{aligned} q(L \mid a) &= \text{pr}(X_{a+1} = 1, Y_{a+1} = 0, X_a = 0 \mid Y_a = 0) + \text{pr}(X_{a+1} = 1, Y_{a+1} = 0, X_a = 1 \mid Y_a = 0) \\ &= q(S \mid a) + q(F0 \mid a) \quad \text{under Assumption 1.} \end{aligned} \tag{A.2}$$

Solving for $q(F0 \mid a)$ in (A.2) and substituting into (A.1) gives $q(N \mid a) = q(F \mid a) + q(I \mid a) + q(S \mid a) - q(L \mid a)$, which can be derived schematically, as in Figure 1. Because we cannot directly estimate $q(L \mid a)$ from subjects who receive a first screening at age a, we eliminate $q(L \mid a)$ using

$$\begin{aligned} q(L \mid a) &= \text{pr}(X_{a+1} = 1 \mid Y_{a+1} = 0, Y_a = 0)\text{pr}(Y_{a+1} = 0 \mid Y_a = 0) \\ &= q(F \mid a+1)(1 - q(N \mid a)), \end{aligned} \tag{A.3}$$

where $q(F \mid a+1) = \text{pr}(X_{a+1} = 1 \mid Y_{a+1} = 0)$ is the probability of a positive first screening at age $a+1$ among subjects age a at the start of the study. Substituting (A.3) into (A.2) gives $q(N \mid a) = q(F \mid a) + q(I \mid a) + q(S \mid a) - q(F \mid a+1)/(1 - q(F \mid a+1))$, which, for a rare disease, is approximately $q(N \mid a) = q(F \mid a) + q(I \mid a) + q(S \mid a) - q(F \mid a+1)$.

To demonstrate why data from multiple screenings can be used to estimate $q(I \mid a)$ and $q(S \mid a)$, let I_j denote a negative screening j at age $a+j$ and symptomatic cancer in the subsequent interval, given negative screening $j-1$ and the subject asymptomatic at age $a+j$. Under the model, the age-specific incidence of I_j is

$$\begin{aligned} &q(I_j \mid a+j) \\ &\quad = \text{pr}(Y_{a+j+1} = 1, X_{a+j} = 0 \mid Y_{a+j} = 0, X_{a+j-1} = 0) \\ &\quad = \text{pr}(Y_{a+j+1} = 1, X_{a+j} = 0, X_{a+j-1} = 0 \mid Y_{a+j} = 0)/\text{pr}(X_{a+j-1} = 0 \mid Y_{a+j} = 0) \\ &\quad = \text{pr}(Y_{a+j+1} = 1, X_{a+j} = 0 \mid Y_{a+j} = 0)/\text{pr}(X_{a+j-1} = 0 \mid Y_{a+j} = 0) \qquad \text{by Assumption 1} \\ &\quad \approx \text{pr}(Y_{a+j+1} = 1, X_{a+j} = 0 \mid Y_{a+j} = 0) \quad \text{because breast cancer is rare} \\ &\quad \approx q(I \mid a+j) \quad \text{by definition.} \end{aligned}$$

Therefore, data from type I detection following the jth screening can used to estimate the age-specific probability of type I detection. We define S_j similarly and apply the same argument.

APPENDIX II

Asymptotic Variance

To compute the asymptotic variance of g, the reduction in cancer mortality due to starting periodic screening at age a instead of age $a+k$, it is convenient to write

$$\begin{aligned} g &= \left(p_a + \sum_{j=1}^{k} \hat{q}(F \mid a+j) m(F, a+j) \right) - \left(p_{a+k} + \sum_{j=0}^{k-1} \hat{q}(F \mid a+j+1) m(F, a+j+1) \right) \\ &= \sum_{j=0}^{k} \hat{q}(F \mid a+j) m(F, a+j) - \sum_{j=0}^{k} \hat{q}(F \mid a+j+1) m(F, a+j+1) \\ &\quad + \sum_{j=0}^{k} \hat{q}(I \mid a+j) m(I, a+j) + \sum_{j=0}^{k} \hat{q}(S \mid a+j) m(S, a+j) \\ &\quad - \sum_{j=0}^{k} \hat{q}(N \mid a+j) m(N, a+j). \end{aligned} \tag{A.4}$$

For the nonrandomized and type I randomized analyses, we can substitute $q(N \mid a) = q(F \mid a) + q(I \mid a) + q(S \mid a) - q(F \mid a+1)$ into (A.4) to obtain

$$\begin{aligned} g &= \sum_{j=0}^{k} \hat{q}(F \mid a+j) m^*(F, a+j) - \sum_{j=0}^{k} \hat{q}(F \mid a+j+1) m^*(F, a+j+1) \\ &\quad + \sum_{j=0}^{k} \hat{q}(I \mid a+j) m^*(I, a+j) + \sum_{j=0}^{k} \hat{q}(S \mid a+j) m^*(S, a+j), \end{aligned} \tag{A.5}$$

where $m^*(d, a+j) = m(d, a+j) - m(N, a+j)$ and $m^*(F, a+j+1) = m(F, a+j+1) - m(N, a+j)$.

For each type of analysis, we compute the asymptotic variance of g using the delta method. Let $\boldsymbol{\lambda}_d = \{\lambda_{dt}\}$ denote the hazards for cancer mortality following type d detection at time t, and let $\boldsymbol{\beta}_d = \{\beta_{di}\}$ denote the parameters modeling the age-specific incidence of type d detection. The variance of g requires $\text{var}(\hat{\boldsymbol{\beta}}_d)$, which is computed from the Poisson regression, and $\text{var}(\hat{\boldsymbol{\lambda}}_d)$, which is a matrix with diagonal elements $\{\lambda_{jd}(1-\lambda_{jd})/n_{jd}\}$, where n_{jd} is the number at risk at the start of time interval j following type d detection. The variance of g also requires $\partial q(d \mid a; \boldsymbol{\beta}_d)/\partial \boldsymbol{\beta}_d + q(d \mid a)(1, a)$ for $q(d \mid a) = \exp(\beta_{0d} + \beta_{1d} a)$ and $\partial m(d, a)/\partial \boldsymbol{\lambda}_{dj} = \Sigma_t \, \partial m(d, a, t)/\partial \boldsymbol{\lambda}_{dj}$, where $m(d, a, t) = surv(a, u) \prod_{j=1}^{u-1} (1 - \hat{\lambda}_{dj}) \hat{\lambda}_{du}$ and $\partial m(d, a, t)/\partial \boldsymbol{\lambda}_{dj} = 0$ if $t < j$, $m(d, a, t)/\lambda_{dj}$ if $t = j$, and $m(d, a, t)/(1 - \lambda_{dj})$ if $t > j$.

Nonrandomized analysis. The asymptotic variance of g is

$$\text{var}(g) = \sum_{d=F,I,S} \left(\frac{\partial g}{\partial \boldsymbol{\beta}_d} \right)^{\mathrm{T}} \text{var}\left(\hat{\boldsymbol{\beta}}_d\right) \left(\frac{\partial g}{\partial \boldsymbol{\beta}_d} \right) + \sum_{d=F,I,S,R} \left(\frac{\partial g}{\partial \boldsymbol{\lambda}_d} \right)^{\mathrm{T}} \text{var}\left(\hat{\boldsymbol{\lambda}}_d\right) \left(\frac{\partial g}{\partial \boldsymbol{\lambda}_d} \right), \tag{A.6}$$

where $\hat{\boldsymbol{\lambda}}_N = \hat{\boldsymbol{\lambda}}_R$, and from (A.5),

$$\frac{\partial g}{\partial \boldsymbol{\beta}_F} = \sum_{j=0}^{k} \frac{\partial q(F \mid a+j)}{\partial \boldsymbol{\beta}_F} m^*(F, a+j) - \sum_{j=0}^{k} \frac{\partial q(F \mid a+j+1)}{\partial \boldsymbol{\beta}_F} m^*(F, a+j+1), \tag{A.7}$$

$$\frac{\partial g}{\partial \boldsymbol{\beta}_d} = \sum_{j=0}^{k} \frac{\partial q(d \mid a+j)}{\partial \boldsymbol{\beta}_d} m^*(d, u), \quad \text{for } d = I, S, \tag{A.8}$$

and, using (A.4) with $q(N \mid a) = q(F \mid a) + q(I \mid a) + q(S \mid a) - q(F \mid a+1)$,

$$\frac{\partial q}{\partial \boldsymbol{\lambda}_F} = \sum_{j=0}^{k} q(F \mid a+j) \frac{\partial m(F, u)}{\partial \boldsymbol{\lambda}_F} - \sum_{j=0}^{k} q(F \mid a+j+1) \frac{\partial m(F, a+j+1)}{\partial \boldsymbol{\lambda}_F}, \tag{A.9}$$

$$\frac{\partial q}{\partial \boldsymbol{\lambda}_d} = \sum_{j=0}^{k} q(d \mid a+j) \frac{\partial m(d,u)}{\partial \boldsymbol{\lambda}_d}, \quad \text{for } d = I, S, \tag{A.10}$$

$$\frac{\partial q}{\partial \boldsymbol{\lambda}_N} = \sum_{j=0}^{k} q(N \mid a+j) \frac{\partial m(N,u)}{\partial \boldsymbol{\lambda}_N}(-1). \tag{A.11}$$

Type I randomized analysis. Using (A.5) with $\hat{\lambda}_{Nt} \approx [\hat{\lambda}_{Ct} - (1-\gamma)\hat{\lambda}_{Rt}]/\gamma$ and treating γ as known, the asymptotic variance of g is

$$\text{var}(g) = \sum_{d=F,I,S} \left(\frac{\partial g}{\partial \boldsymbol{\beta}_d}\right)^{\mathrm{T}} \text{var}\left(\hat{\boldsymbol{\beta}}_d\right) \left(\frac{\partial g}{\partial \boldsymbol{\beta}_d}\right) + \sum_{d=F,I,S,C,R} \left(\frac{\partial g}{\partial \boldsymbol{\lambda}_d}\right)^{\mathrm{T}} \text{var}\left(\hat{\boldsymbol{\lambda}}_d\right) \left(\frac{\partial g}{\partial \boldsymbol{\lambda}_d}\right),$$

with (A.7), (A.8), (A.9), (A.10), (A.11),

$$\frac{\partial g}{\partial \boldsymbol{\lambda}_C} = \frac{\partial g}{\partial \boldsymbol{\lambda}_N} \frac{1}{\gamma},$$

and

$$\frac{\partial g}{\partial \boldsymbol{\lambda}_C} = -\frac{\partial g}{\partial \boldsymbol{\lambda}_N} \frac{1-\gamma}{\gamma}.$$

Type II randomized analysis. Treating π and γ as known, the asymptotic variance is

$$\text{var}(g) = \sum_{d=F,I,S,R,C} \left(\frac{\partial g}{\partial \boldsymbol{\beta}_d}\right)^{\mathrm{T}} \text{var}\left(\hat{\boldsymbol{\beta}}_d\right) \left(\frac{\partial g}{\partial \boldsymbol{\beta}_d}\right) + \sum_{d=F,I,S,C,R} \left(\frac{\partial g}{\partial \boldsymbol{\lambda}_d}\right)^{\mathrm{T}} \text{var}\left(\hat{\boldsymbol{\lambda}}_d\right) \left(\frac{\partial g}{\partial \boldsymbol{\lambda}_d}\right),$$

where, from (A.4), with $\hat{\lambda}_{Nt} \approx [\hat{\lambda}_{Ct} - (1-\gamma)\hat{\lambda}_{Rt}]/\gamma$ and $\hat{q}(N \mid a+j) = \hat{q}(C \mid a+j)(1-\pi)/\pi - \hat{q}(N \mid a+j)/\pi$,

$$\frac{\partial g}{\partial \boldsymbol{\beta}_F} = \sum_{j=0}^{k} \frac{\partial q(F \mid a+j)}{\partial \boldsymbol{\beta}_F} m(F, a+j) - \sum_{j=0}^{k} \frac{\partial q(F \mid a+j+1)}{\partial \boldsymbol{\beta}_F} m(F, a+j+1),$$

$$\frac{\partial g}{\partial \boldsymbol{\beta}_d} = \sum_{j=0}^{k} \frac{\partial q(d \mid a+j)}{\partial \boldsymbol{\beta}_d} m(d,u), \quad \text{for } d = I, S,$$

$$\frac{\partial g}{\partial \boldsymbol{\beta}_C} = \sum_{j=0}^{k} \frac{\partial q(C \mid a+j)}{\partial \boldsymbol{\beta}_C} \frac{1}{\pi} m(N,u),$$

$$\frac{\partial g}{\partial \boldsymbol{\beta}_R} = -\sum_{j=0}^{k} \frac{\partial q(R \mid a+j)}{\partial \boldsymbol{\beta}_R} \frac{(1-\pi)}{\pi} m(N,u),$$

(A.9), (A.10), and (A.11).