



Practice of Epidemiology

Correcting for Lead Time and Length Bias in Estimating the Effect of Screen Detection on Cancer Survival

Stephen W. Duffy¹, Iris D. Nagtegaal², Matthew Wallis³, Fay H. Cafferty¹, Nehmat Houssami⁴, Jane Warwick¹, Prue C. Allgood¹, Olive Kearins⁵, Nancy Tappenden⁵, Emma O'Sullivan⁵, and Gill Lawrence⁵

¹ Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute for Preventive Medicine, London, United Kingdom.

² Department of Pathology, Dutch Cancer Society, Nijmegen, the Netherlands.

³ Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge, United Kingdom.

⁴ Screening and Test Evaluation Program, School of Public Health, University of Sydney, Sydney, Australia.

⁵ West Midlands Breast Screening Quality Assurance Reference Centre, University of Birmingham, Birmingham, United Kingdom.

Received for publication December 21, 2007; accepted for publication April 8, 2008.

Determination of survival time among persons with screen-detected cancer is subject to lead time and length biases. The authors propose a simple correction for lead time, assuming an exponential distribution of the pre-clinical screen-detectable period. Assuming two latent categories of tumors, one of which is more prone to screen detection and correspondingly less prone to death from the cancer in question, the authors have developed a strategy of sensitivity analysis for various magnitudes of length bias. Here they demonstrate these methods using a series of 25,962 breast cancer cases (1988–2004) from the West Midlands, United Kingdom.

bias (epidemiology); breast neoplasms; mass screening; models, statistical; survival

Abbreviation: CI, confidence interval.

In disease screening, the concepts of lead time and length bias have been familiar for decades (1, 2). Lead time is the amount of time by which the diagnosis has been advanced by screening. In analysis of survival from diagnosis, it constitutes an artificial addition to the survival time of screen-detected cases. Length bias is the phenomenon whereby more slowly growing tumors, with less capacity to prove fatal, may have a longer presymptomatic screen-detectable period and will therefore be more likely to be screen-detected. This again confers an artificial survival advantage to screen-detected cases. The extreme form of length bias is overdiagnosis, defined as diagnosis by screening of cancers which would not have come to clinical attention in the host's lifetime had screening not taken place. It is thought, for

example, that some in situ cancers (cancers confined to the ducts which have not yet invaded the surrounding tissue) detected by screening might never have become invasive or given rise to symptoms in the absence of screening (3).

To avoid these biases, investigators in randomized trials of cancer screening compare mortality rates from the disease in question in the whole population randomized to screening with those in the whole control population, instead of comparing survival rates of disease cases. In addition, the time origin is taken as the point of randomization, not the point of diagnosis (4, 5). However, the effect of screen detection on case survival is often of interest, particularly in the case of mammographic screening for breast cancer, which is now in the post-trial epoch. The emphasis

Correspondence to Dr. Prue Allgood, Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute for Preventive Medicine, Charterhouse Square, London WC1M 6BQ, United Kingdom (e-mail: prue.allgood@cancer.org.uk).

now is on evaluation of routine screening services and on assessing screening programs in special risk groups for which randomized trials may not be feasible or ethical (6).

We have developed a simple method of correction for lead time in analysis of survival including screen-detected cases and an approach to sensitivity analyses for length bias. The lead-time method owes much to the earlier work of Walter and Stitt (7) but is rather simpler and less general than their method. The length-bias approach involves assuming two latent tumor populations, one with both a higher probability than the other of being screen-detected and a correspondingly lower probability of fatality, whether symptomatic or screen-detected. Note that these methods are not tools with which to evaluate the efficacy of cancer screening. The appropriate method for determining whether a cancer screening strategy works is the randomized controlled trial, with mortality as the endpoint. Here we apply these methods to results from a large series of breast cancer cases from the West Midlands, United Kingdom.

METHODS

Correction for lead-time bias

Correction for lead-time bias involves estimation of the additional follow-up time observed purely as a result of lead time in a case of screen-detected cancer. We assume an exponential distribution of the sojourn time (8), the period during which the tumor is asymptomatic but screen-detectable, with a rate of transition to symptomatic disease λ . Thus, $1/\lambda$ is the mean sojourn time. Consider first a screen-detected cancer which has resulted in breast cancer death at time t after diagnosis. The additional follow-up time cannot be greater than t . The expected additional follow-up time, s , due to lead time is the expectation of the lead time conditional on its being less than t , that is,

$$E(s) = \frac{\int_0^t x e^{-\lambda x} dx}{1 - e^{-\lambda t}} = \frac{1 - e^{-\lambda t} - \lambda t e^{-\lambda t}}{\lambda(1 - e^{-\lambda t})}. \tag{1}$$

For a patient with a screen-detected cancer known to be alive at time t after diagnosis, we have

$$E(s) = P(s \leq t)E(s | s \leq t) + P(s > t)t.$$

So in this case,

$$E(s) = \frac{(1 - e^{-\lambda t}) \int_0^t x e^{-\lambda x} dx}{1 - e^{-\lambda t}} + t e^{-\lambda t} = \frac{1 - e^{-\lambda t}}{\lambda}. \tag{2}$$

Thus, a correction for lead time would involve subtracting $E(s)$ from the observed survival time, or time to last live follow-up, of patients with screen-detected cases. By expansion of the exponential terms, it can be seen that $E(s)$ is never greater than t , so the correction never gives negative survival times.

Sensitivity analyses to adjust for length bias

Length-biased sampling occurs when the chance of an observation's being in a sample is proportional to a particular

characteristic of the observation. In the context of screening, "length bias" is used to refer to the phenomenon whereby slower-growing, less aggressive tumors have a longer preclinical screen-detectable period and are therefore more likely to be screen-detected than faster-growing, more aggressive cancers. Let us assume that tumors can fall into two categories, one with a probability p_s of being screen-detected (category A) and the other with a probability p_s/θ of being screen-detected, where $\theta < 1$ and $p_s/\theta < 1$ (category B). Suppose that a proportion q of the tumors is in category A and the complementary proportion $1 - q$ is in category B. The probabilities of being screen-detected will depend on the screening regimen offered in terms of frequency and sensitivity and the rate of participation in screening. We assume these to be uniform within the particular screening program under study.

Category B tumors have a greater chance of being screen-detected. If we assume that this is because the tumors are slow-growing (this group might include a large proportion of in situ cases), it seems reasonable to assume that they are correspondingly less likely to cause death. Suppose, therefore, that the probability of death from breast cancer during the period of observation from a category A symptomatic tumor is p and the corresponding probability for category B symptomatic cancers is θp . Thus, the tumors more likely to be screen-detected are postulated to be correspondingly less likely to cause death a priori—the classic manifestation of length bias in a screening context.

Assume further that the true relative risk of death from the cancer in question for screen-detected versus symptomatic tumors in a population offered screening, independent of length bias, is ϕ . This means that within each category, screen-detected cancers are ϕ times as likely to cause death as symptomatic tumors. Correspondingly, we assume that within each detection mode (screening and symptomatic), patients with category B tumors are θ times as likely to die from breast cancer as patients with category A tumors. Thus, death rates from breast cancer in the four categories are as follows:

- Category A symptomatic tumors: p .
- Category B symptomatic tumors: θp .
- Category A screen-detected tumors: ϕp .
- Category B screen-detected tumors: $\theta \phi p$.

The value of ϕ will depend on characteristics of the symptomatic tumors as well as on the effectiveness of the screening. Symptomatic tumors will typically include both interval cancers (cancers arising symptomatically among screening participants in the intervals between screens) and cancers diagnosed symptomatically in women who chose not to attend screening (for brevity we shall refer to these women as nonattenders). If, as has been observed in the past, cancers arising in nonattenders have particularly poor outcomes, ϕ will depend on the proportion of nonattender cancers among the symptomatic tumors, and therefore on the attendance rate in the screening program under study.

The observed probability of cancer death for symptomatic tumors in a population offered screening will be

$$\begin{aligned}
 p(D|\text{sympt}) &= p_1 \\
 &= \frac{\{q(1-p_s)p + (1-q)(1-\frac{p_s}{\theta})\theta p\}}{q(1-p_s) + (1-q)(1-\frac{p_s}{\theta})} \\
 &= \frac{\theta p(q-p_s q + \theta - q\theta - p_s + qp_s)}{\theta - qp_s\theta - (1-q)p_s} \\
 &= \frac{\theta p(q(1-\theta) + (\theta - p_s))}{\theta - qp_s\theta - (1-q)p_s}.
 \end{aligned}$$

The probability of cancer death for screen-detected tumors in the same population will be observed to be

$$\begin{aligned}
 p(D|\text{asympt}) &= p_2 \\
 &= \frac{(qp_s\phi p + (1-q)p_s\phi p)}{qp_s + (1-q)\frac{p_s}{\theta}} \\
 &= \frac{p_s\phi p\theta}{qp_s\theta + (1-q)p_s} \\
 &= \frac{\phi p\theta}{q\theta + 1 - q}.
 \end{aligned}$$

The observed relative risk of cancer death will therefore be

$$\begin{aligned}
 R &= \frac{p_2}{p_1} \\
 &= \frac{\phi p\theta}{q\theta + 1 - q} \bigg/ \frac{\theta p(q(1-\theta) + (\theta - p_s))}{\theta - qp_s\theta - (1-q)p_s} \\
 &= \frac{\phi(\theta - qp_s\theta - (1-q)p_s)}{(q\theta + 1 - q)(q(1-\theta) + (\theta - p_s))}.
 \end{aligned}$$

This will be smaller than ϕ , provided that q and θ are less than unity, so it will overestimate the reduction conferred by screen detection.

The overall probability of screen detection is

$$\begin{aligned}
 p(\text{asympt}) &= p_3 \\
 &= qp_s + (1-q)\frac{p_s}{\theta} \\
 &= \frac{p_s}{\theta}(\theta q + 1 - q).
 \end{aligned}$$

We can estimate p_s , p , and ϕ in terms of p_1 , p_2 , p_3 , θ , and q , as follows:

$$\begin{aligned}
 \hat{p}_s &= \frac{p_3\theta}{(\theta q + 1 - q)}; \\
 \hat{p} &= \frac{p_1(\theta - qp_s\theta - (1-q)p_s)}{\theta(q(1-\theta) + (\theta - p_s))}. \tag{3}
 \end{aligned}$$

After substitution of p_s and some algebra, we have

$$\hat{p} = \frac{p_1(\theta q + 1 - q)(1 - p_3)}{(\theta q + 1 - q)(\theta + q(1 - \theta)) - p_3\theta}. \tag{4}$$

Finally,

$$\begin{aligned}
 \hat{\phi} &= \frac{p_2(q\theta + 1 - q)}{\theta p} \\
 &= \frac{p_2\{(\theta q + 1 - q)(\theta + q(1 - \theta)) - p_3\theta\}}{p_1\theta(1 - p_3)}. \tag{5}
 \end{aligned}$$

We can estimate p_1 , p_2 , and p_3 from screening data. A range of assumed values can be posited for θ and q to obtain a likely range of true values for ϕ .

An estimate of the relative hazard assuming exponential survival is

$$\hat{\psi} = \frac{\ln(1 - \hat{\phi}\hat{p})}{\ln(1 - \hat{p})}.$$

EXAMPLE

In collaboration with breast cancer screening units, the West Midlands Cancer Intelligence Unit collected clinico-pathologic, diagnostic, and follow-up data on cancers diagnosed in the county of West Midlands, United Kingdom, among women aged 50–69 years from 1988 to 2001 and among women aged 50–74 years from 2002 to 2004. We had data on 10,100 screen-detected breast cancers and 15,862 symptomatic breast cancers (6,009 interval cancers and 9,853 tumors in nonattenders). Thus, p_3 , the observed probability of screen detection, was 0.39. There were 4,935 deaths among the symptomatic cases (4,620 within 10 years of diagnosis) and 929 deaths among the screen-detected cases (819 within 10 years). The observed 10-year case fatality for symptomatic tumors was $p_1 = 0.35$, and for screen-detected tumors it was $p_2 = 0.12$. This produces a relative risk of 0.34 (95 percent confidence interval (CI): 0.31, 0.37). The corresponding relative hazard from Cox regression is 0.27 (95 percent CI: 0.25, 0.30).

First, we correct for lead time. In the Swedish Two-County Trial, a study of breast screening, Tabar et al. (9) estimated λ as 0.27 for the age group 50–59 years and 0.24 for the age group 60–69 years, with a weighted average of 0.25. Figure 1 shows the uncorrected and lead-time-corrected survival for the screen-detected cancers as compared with symptomatic cancers. The survival of symptomatic cases is unchanged, but the correction has led to a lower survival estimate in the screen-detected cases. After the correction, there were 906 breast cancer deaths within 10 years among the screen-detected cases. The 10-year case fatality for the screen-detected cases, corrected for lead time, was 0.17. The relative risk was 0.49 (95 percent CI: 0.45, 0.53), and the Cox regression relative hazard was 0.40 (95 percent CI: 0.37, 0.44).

For the correction for length bias, we do not know the values of q and θ , so we calculate the corrected results for a range of plausible values. Results are shown in table 1. The correction is more dependent on θ , the relative rate of screen detection and fatality in the length-bias group, than on q , the complement of the group's size. This range of values yields estimates of the true relative risk ranging from 0.49 to 0.59, with a median of 0.51, and estimates of the relative hazard

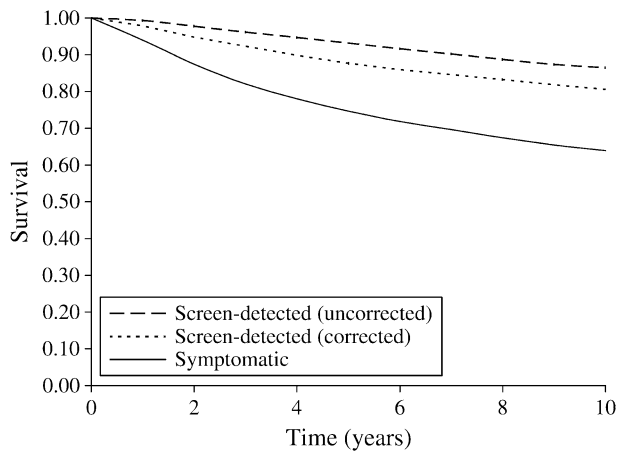


FIGURE 1. Survival of women with symptomatic and screen-detected cases of breast cancer, before and after correction for lead time, West Midlands, United Kingdom, 1988–2004.

ranging from 0.43 to 0.52, with a median of 0.45. Further analyses showed that for the length bias to account for the entire difference in survival, we would require $q = 0.7$ and $\theta = 0.2$; that is, the length-bias group would have to comprise at least 30 percent of the tumor population and be five times more likely to be screen-detected and five times less likely, a priori, to cause death from breast cancer.

The above analysis is based on all tumors in the series. Excluding the 2,102 cases of in situ carcinoma, the uncorrected 10-year survival in the screen-detected cases was 86 percent and that in the symptomatic cases was 64 percent, a relative risk of 0.39 (95 percent CI: 0.36, 0.42). The Cox regression relative hazard was 0.31 (95 percent CI: 0.28, 0.33). Correcting for lead time, the 10-year survival in the screen-detected cases was 81 percent, giving a relative risk of 0.53 (95 percent CI: 0.49, 0.57); the relative hazard was 0.45 (95 percent CI: 0.42, 0.49). Sensitivity analysis for length bias using the range of values of q and θ in table 1 gave a range of values for the relative risk from 0.53 to 0.63, with a median of 0.55. The relative hazard ranged from 0.47 to 0.56, with a median of 0.49. The results of both lead-time and length-bias corrections are summarized in table 2.

DISCUSSION

The above work demonstrates a simple correction for lead-time bias in analysis of cancer survival data involving screen-detected cases, as well as a relatively simple approach to sensitivity analysis for length bias. A STATA routine for the lead-time correction and an Excel spreadsheet for the length-bias sensitivity analyses are available from the authors. The modeling requires strong assumptions but has the advantage of being easy to carry out. Part of the simplicity of the length-bias analysis lies in the fact that we have modeled it as an effect on relative risk, the ratio

of probabilities of dying of the disease within a specified time, rather than an effect on relative hazard, the ratio of the instantaneous rates of death. This means that the relative hazards calculated in the length-bias analyses are approximate, being dependent on the absolute probabilities of dying and therefore on the period of observation. In addition, there is the assumption that in the length-bias population (category B), the proportional increase in screen detection propensity is equal to the proportional decrease in cause-specific fatality. Length bias is a well-known phenomenon, but it is difficult to estimate or otherwise quantify. The method demonstrated here at least provides a means of estimating a likely range for its effect.

We reiterate that the methods shown here are not intended for use in establishing whether or not screening works in principle. This is determined by randomized trials with mortality as the endpoint. It is also worth bearing in mind that the relative risk estimated here is for screen-detected cases as compared with symptomatic cases. In a population randomized controlled trial, the mortality relative risk for the study group offered screening as compared with a control group not offered screening would dilute the relative risk estimated here by the cancers in the study arm diagnosed in nonattenders or as interval cancers. The expected dilution might be around 35 percent (10). Thus, the corrected 45 percent reduction in risk ($\phi = 0.55$, above) would translate to 29 percent (0.65×45 percent)—a relative risk of 0.71, which is compatible with trial results (4, 10). The corrected relative hazard of 0.55 is similar to the 0.57 observed for first screen-detected cancers compared with control symptomatic cancers in the Swedish Two-County Trial, adjusting for size, lymph node status, and grade, which would tend to adjust out the effects of lead time and length bias, in addition to some of the genuine survival advantage of screen detection.

Note that there are other biases in comparing survival among screen-detected cases with that among symptomatic cases, particularly the fact that outside of the randomized trial setting, there may be healthy volunteer bias in that the symptomatic cases may include persons who have declined the offer of screening. This population may be less health-conscious and therefore more likely to die of their disease regardless of screening. This can be addressed by comparing screen-detected cancers with interval cancers, those which arise in screening participants but symptomatically in the intervals between screens, or by using published methods for correction for the healthy volunteer bias (10).

It is interesting that in the example considered, the lead-time correction makes a substantial difference in the estimated relative risk of dying from breast cancer, but the length-bias adjustment makes a smaller difference. Note that the length-bias correction is dependent primarily on the values of θ and q and is not affected by the fact that the lead-time correction was carried out first. To make a difference which would correct the relative risk to unity in this example would require length bias of an implausible magnitude. In this method, we have treated length bias and lead time as two independent phenomena, but in fact they are mutually associated, since a length-bias case will have a longer sojourn time and potentially, therefore, a longer lead time. Our method adjusts for the expected lead time based

TABLE 1. Estimates of p_s , p , and ϕ for varying values of θ and q , for lead-time-corrected fatality rates of 0.17 and 0.35, and for an observed relative risk of 0.49 among screen-detected and symptomatic cases of breast cancer, West Midlands, United Kingdom, 1988–2004

Relative screen detection and fatality rate, θ	Mixing fraction, q	Category A screen detection probability, p_s	Category A symptomatic tumor fatality rate, p	True relative risk, ϕ	Approximate relative hazard
0.9	0.9	0.39	0.35	0.49	0.43
0.9	0.8	0.38	0.36	0.49	0.43
0.9	0.7	0.38	0.36	0.49	0.43
0.9	0.6	0.37	0.36	0.49	0.43
0.9	0.5	0.37	0.37	0.49	0.43
0.8	0.9	0.38	0.36	0.49	0.43
0.8	0.8	0.37	0.36	0.49	0.44
0.8	0.7	0.36	0.37	0.49	0.44
0.8	0.6	0.35	0.38	0.50	0.44
0.8	0.5	0.35	0.39	0.50	0.44
0.7	0.9	0.37	0.36	0.49	0.44
0.7	0.8	0.36	0.37	0.50	0.45
0.7	0.7	0.35	0.38	0.51	0.45
0.7	0.6	0.33	0.39	0.51	0.45
0.7	0.5	0.32	0.40	0.51	0.45
0.6	0.9	0.37	0.36	0.50	0.45
0.6	0.8	0.34	0.37	0.52	0.46
0.6	0.7	0.33	0.38	0.53	0.47
0.6	0.6	0.31	0.40	0.54	0.47
0.6	0.5	0.29	0.42	0.54	0.47
0.5	0.9	0.35	0.36	0.52	0.47
0.5	0.8	0.33	0.37	0.55	0.49
0.5	0.7	0.30	0.39	0.57	0.51
0.5	0.6	0.28	0.41	0.58	0.52
0.5	0.5	0.26	0.44	0.59	0.51

on the average sojourn time first, and then carries out a series of sensitivity analyses for different magnitudes of heterogeneity around the average sojourn time.

The mean sojourn estimate which we used for the lead time correction was 4 years, from the Swedish Two-County Trial (9). Other studies yield similar or smaller estimates (11–14), with only one exception (15), so it is likely that 4 years is accurate and may even be conservative.

We can use algebra similar to that of the length-bias analyses above to estimate the effect of the most extreme form of length bias, overdiagnosis. Suppose that in category B, the tumors would never have become symptomatic and would never have caused death, as is almost certainly the case for some in situ tumors. Then the observed case fatality rate of the symptomatic tumors would be

$$p_1 = \frac{pq(1-p_s)}{q(1-p_s)} = p.$$

The observed proportion of cases that were screen-detected would be

$$p_3 = qp_s + 1 - q.$$

The observed case fatality rate of persons with the screen-detected tumors would be

$$p_2 = \frac{qp_s\phi p}{qp_s + 1 - q}.$$

This would give

$$\phi = \frac{R(qp_s + 1 - q)}{qp_s}.$$

Our estimates of overdiagnosis tend to be small, on the order of 10 percent or less (13), but we assume a more extreme case for demonstration purposes. If we assume that 25 percent of screen-detected cases are overdiagnosed, $p = 0.39$ gives

$$0.25 = \frac{1 - q}{0.39},$$

which gives $q = 0.9$. This in turn gives $p_s = 0.32$ and $\phi = 0.66$.

TABLE 2. Results of survival analysis, both uncorrected and corrected for lead time and length bias, among screen-detected and symptomatic cases of breast cancer, West Midlands, United Kingdom, 1988–2004

Parameter measured	Symptomatic breast cancers	Screen-detected breast cancers	Relative risk	95% CI* or range	Cox regression relative hazard	
					Hazard ratio	95% CI or range
Total no. of breast cancers	15,862	10,100	NA*		NA	
No. of in situ breast cancers	535	1,567				
No. of microinvasive and invasive breast cancers	15,327	8,533				
Total no. of deaths	4,935	929				
Total no. of deaths at 10 years	4,620	819				
Total no. of deaths at 10 years after correction for lead time	4,620	906				
Total no. of in situ breast cancer deaths at 10 years	3	1				
Total no. of microinvasive and invasive breast cancer deaths at 10 years	4,617	818				
10-year case fatality (all cases)	0.35	0.12	0.34	0.31, 0.37	0.27	0.25, 0.30
10-year survival (all cases) (%)	65	88				
10-year case fatality (excluding in situ cases)	0.36	0.14	0.39	0.36, 0.42	0.31	0.28, 0.33
10-year survival (excluding in situ cases) (%)	64	86				
Results obtained after application of lead-time correction						
10-year case fatality (all cases)	0.35	0.17	0.49	0.45, 0.53	0.40	0.37, 0.44
10-year survival (all cases) (%)	65	83				
10-year case fatality (excluding in situ cases)	0.36	0.19	0.53	0.49, 0.57	0.45	0.42, 0.49
10-year survival (excluding in situ cases) (%)	64	81				
Results obtained after additional application of length-bias correction						
10-year case fatality (all cases)	0.35	0.18	0.51	0.49–0.59†	0.45	0.43–0.52†
10-year survival (all cases) (%)	65	82				
10-year case fatality (excluding in situ cases)	0.36	0.20	0.55	0.53–0.63†	0.49	0.47–0.56†
10-year survival (excluding in situ cases) (%)	64	80				

* CI, confidence interval; NA, not applicable.

† For length-bias adjustment, absolute ranges of values from sensitivity analyses are given instead of 95% confidence intervals.

The lead time correction is applied as a constant, so the 95 percent confidence intervals will be slightly anticonservative, since they do not reflect the additional uncertainty from estimation of λ . However, the correction is applied to a minority of observations, the variance of our estimate of λ is small (9), and with large data sets such as this, the confidence intervals would be narrow even if they did incorporate uncertainty in estimation of λ .

The corrections demonstrated in this paper are specific to the screening program providing the data. In the example shown, they pertain to the United Kingdom program, which throughout the period of observation was mainly using two-view mammography every 3 years. A program with more intensive or more frequent screening would have a different probability of screen detection and therefore different correction factors. In addition, the lead-time correction applied is based on the average sojourn time over all cancers. It could be argued that a better correction would be based on each individual tumor's stage at diagnosis. This, however, would involve considerable algebraic and analytic complexity.

The effect of the policy of screening is best evaluated using population mortality from the disease in question. However, the effect of screen detection on case survival is of considerable interest to the clinicians treating cancer patients and to the patients themselves. The work above provides a means of estimating survival rates taking into account the major biases inherent in such estimates.

ACKNOWLEDGMENTS

Dr. Prue Allgood was supported by a grant from the Princess Grace Hospital, London. Dr. Iris Nagtegaal was supported by a fellowship from the Dutch Cancer Society. The West Midlands Screening Histories Project was supported by a grant from the Breast Cancer Research Trust.

The authors are grateful to the National Health Service Trusts, private hospitals, and National Health Service breast screening services in the West Midlands for providing

cancer registration and breast screening data. They are also grateful to Rosie Day of the West Midlands Cancer Intelligence Unit for extracting the breast cancer data from the Unit's cancer registration database.

Conflict of interest: none declared.

REFERENCES

1. Hutchison GB, Shapiro S. Lead time gained by diagnostic screening for breast cancer. *J Natl Cancer Inst* 1968;41:665–81.
2. Zelen M, Fainleib M. On the theory of screening for chronic disease. *Biometrika* 1969;56:601–14.
3. Yen MF, Tabar L, Vitak B, et al. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer* 2003;39:1746–54.
4. 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.
5. Hardcastle JD, Chamberlain JO, Michael HE, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996;348:1472–7.
6. Maurice A, Evans DGR, Shenton A, et al. Screening younger women with a family history of breast cancer—does early detection improve outcome? *Eur J Cancer* 2006;42:1385–90.
7. Walter SD, Stitt LW. Evaluating the survival of cancer cases detected by screening. *Stat Med* 1987;6:885–900.
8. Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures for mass screening programs. *Biometrics* 1984;40:1–14.
9. Tabar L, Vitak B, Chen HH, et al. The Swedish Two-County Trial twenty years later: updated mortality results and new insights from long term follow-up. *Radiol Clin North Am* 2000;38:625–51.
10. Duffy SW, Cuzick J, Tabar L, et al. Correcting for non-compliance bias in case-control studies to evaluate cancer screening programs. *Appl Stat* 2002;51:235–43.
11. Paci E, Duffy SW. Modelling the analysis of breast cancer screening programmes: sensitivity, lead time and predictive value in the Florence District Programme (1975–1986). *Int J Epidemiol* 1991;20:852–8.
12. van Oortmarssen GJ, Habbema JD, van der Maas PJ, et al. A model for breast cancer screening. *Cancer* 1990;66:1601–12.
13. Cong XJ, Shen Y, Miller AB. Estimation of age-specific sensitivity and sojourn time in breast cancer screening studies. *Stat Med* 2005;24:3123–38.
14. Olsen AH, Agbaje OF, Myles JP, et al. Overdiagnosis, sojourn time and sensitivity in the Copenhagen mammography screening programme. *Breast J* 2006;12:338–42.
15. Weedon-Fekjaer H, Vatten LJ, Aalen OO, et al. Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening: new results. *J Med Screen* 2005;12:172–8.