DOI: 10.1111/j.1075-122X.2006.00272.x

ORIGINAL ARTICLE

Overdiagnosis, Sojourn Time, and Sensitivity in the Copenhagen Mammography Screening Program

Anne Helene Olsen, PhD,* Olorunsola F. Agbaje, PhD,* Jonathan P. Myles, PhD,* Elsebeth Lynge,† and Stephen W. Duffy, MSc*

**Cancer Research UK, Department of Epidemiology, Mathematics, and Statistics, Wolfson Institute of Preventive Medicine, Charterhouse Square, London, United Kingdom; and †Institute of Public Health, University of Copenhagen, Copenhagen, Denmark*

■ **Abstract:** The goal of this research was to estimate the overdiagnosis at the first and second screens of the mammography screening program in Copenhagen, Denmark. This study involves a mammography service screening program in Copenhagen, Denmark, with 35,123 women screened at least once. We fit multistate models to the screening data, including preclinical incidence of progressive cancers and nonprogressive (i.e., overdiagnosed) cancers. We estimated mean sojourn time as 2.7 years (95% confidence interval [CI] 2.2–3.1) and screening test sensitivity as 100% (95% CI 99.8–100). Overdiagnosis was estimated to be 7.8% (95% CI 0.3–26.5) at the first screen and 0.5% (95% CI 0.02–2.1) at the second screen. This corresponds to 4.8% of all cancers diagnosed among participants during the first two invitation rounds and following intervals. A modest overdiagnosis was estimated for the Copenhagen screening program, deriving almost exclusively from the first screen. The CIs were very broad, however, and estimates from larger datasets are warranted. ■

Key Words: breast, malignant neoplasms, mammography, overdiagnosis, screening

Mammography screening for breast cancer is justifiable only if it reduces breast cancer mortality. This is a necessary but not a sufficient requirement, since mammography screening has potential negative side effects. One of these is the possibility of overdiagnosis, that is, diagnosis of breast cancers that without screening would not have emerged clinically in the woman's lifetime.

A concept often encountered in cancer screening evaluation is length bias. This is the tendency for slow-growing, less aggressive cases to be detected by screening, as they will tend to have long preclinical screen-detectable periods. Overdiagnosis may be thought of as an extreme form of length bias, where the tumor develops so slowly that it would never have given rise to symptoms in the lifetime of the host.

The potential overdiagnosis of invasive breast cancers in the Copenhagen program was studied by analyzing the trends in breast cancer incidence in the period before and after the start of screening, comparing Copenhagen with the nonscreening regions in Denmark (1). The study showed no indication of overdiagnosis in the second to the fifth invitation rounds, where the incidence assumed a level compatible with that expected in the absence of screening.

It is, however, not easy to assess possible overdiagnosis in the first invitation round based on trends. In the period of the first invitation round, breast cancers detected include those that would also have been detected without screening, preclinical screen-detectable breast cancers (earlier diagnosis), and potentially, overdiagnosed breast cancers. This results in a high breast cancer incidence in the first invitation round, often referred to as the prevalence peak. The magnitude of the peak depends on the participation rate in the program, the sensitivity of the screening test, the lead time, and the degree of overdiagnosis.

In this article we attempt to estimate the degree of overdiagnosis at the first and second screens, taking into account the test sensitivity, preclinical incidence of breast cancer, and mean sojourn time based on the first and second invitation rounds of the mammography screening program in Copenhagen (2).

Address correspondence and reprint requests to: Anne Helene Olsen, PhD, Cancer Research UK, Department of Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventive Medicine, Charterhouse Square, London, EC1M 6BQ, UK, or e-mail: stephen.duffy@cancer.org.uk.

MATERIALS AND METHODS

The Copenhagen mammography screening program started in 1991, offering mammography screening biennially to women age 50–69 years at the beginning of each

invitation round. About 40,000 women are invited in each invitation round.

The number of women screened was retrieved from the databases of the mammography screening program in Copenhagen. The final diagnoses of screen-detected cancers (invasive or in situ) were found by linkage of these with the Danish Cancer Registry and the databases of the Danish Breast Cancer Cooperative Group (DBCG). The registries contain data on invasive breast cancers diagnosed in Denmark since 1943 and 1977, respectively. In addition, the DBCG contains data on in situ breast cancers. Inconsistent information was checked in the Pathology Registry.

The number of interval cancers was retrieved by merging the mammography screening databases with the Danish Cancer Registry and the Central Population Registry. Interval cancers were defined as invasive breast cancers diagnosed in women with a negative screening result in the period between the screening date and the date of death, emigration, next screening date, or 2 years since the previous screening, whichever came first. Databases were linked by the unique personal identification number issued to all residents of Denmark.

Data from the first and second invitation rounds were included and tabulated by screen number. Data for women participating in the second invitation round were included as first screen if they had not participated in the first invitation round, and as second screen if they had participated in the first invitation round.

For estimation of overdiagnosis we used models similar to those of Day and Walter (3). We assumed a uniform annual incidence, I, of preclinical but screen-detectable, truly progressive cancers, an exponential distribution of time from inception of these to clinical symptoms with rate λ, and a screening test sensitivity, S. In addition, we assumed an exponential incidence of nonprogressive, and therefore overdiagnosed, preclinical screen-detectable cancers with rate μ. Since a tumor is only overdiagnosed if it is actually detected at screening, we defined the screening test sensitivity to be 100% for overdiagnosed cancers. The expected rates of cancers diagnosed at the first and second screens, and in the intervals following those screens with an average interval time of t are as follows:

First screen:

$$SI/\lambda + (1 - e^{-\mu a}),$$

where a is the average age; 60 years in this case.

Between first and second screen:

$$I/\lambda\{-S(1 - e^{-\lambda t}) + \lambda t\}.$$

Second screen:

$$SI/\lambda(1 - Se^{-\lambda t}) + (1 - e^{\mu t}).$$

Between second and third screen:

$$I/\lambda\{(1 - S)(1 - Se^{-\lambda t})(1 - e^{-\lambda t}) + \lambda t - (1 - e^{-\lambda t})\}.$$

From the data on screen-detected and interval cancers, we estimated I, λ, S, and μ by fitting Poisson distributions to the numbers of cases at the two screens and in the two intervals, with expectations as above. The interscreening interval was 2 years, so $t = 2$. The estimation algorithm used was Markov chain Monte Carlo, implemented in the computer program WinBUGS (4). Prior distributions used for the parameters I, λ, S, and μ were as follows:

I: lognormal (0.0, 0.0001)
λ: gamma (0.1, 1.0)
S: logit(S) = α, $\alpha \sim$ normal (0.0, 0.0001)
μ: uniform (0.0, 1.0).

Note that the second parameter in the normal and lognormal distributions is the precision, not the variance or the standard deviation (5). The major assumptions here are a uniform rate of incidence of progressive preclinical disease, an exponential distribution of time spent in the preclinical state, and an exponential distribution of time to incidence of overdiagnosable, nonprogressive disease, allowing such disease to occur at any point in the host's lifetime. This last assumption may be false, but due to the shape of the exponential distribution it gives expected numbers detected at screening very similar to those resulting from arbitrarily imposing a lower age limit of, for example, 30 years.

RESULTS

Table 1 shows the cancers diagnosed in the first two rounds of screening, by detection mode. In the first and second invitation rounds, 35,123 women had their first screen, and of these there were 379 screen-detected cancers and 67 interval cancers in women screened negative. Of

Table 1. Findings of the Copenhagen Screening Program for Patients Age 50–69 Years

Detection occasion	Number screened	Total cancers	Invasive cancers
First screen	35,123	379	329
First interval	35,123	67	63
Second screen	21,307	123	113
Second interval	21,307	58	53

Table 2. Estimates from Overdiagnosis Modeling

Quantity	Estimate	95% CI
Incidence (true cases)/1000	3.8	3.3–4.2
Screening test sensitivity (%)	100	99.8–100.0
Mean sojourn time (years)	2.7	2.2–3.1
Incidence of overdiagnosed cases/1000	0.0142	0.0005–0.0493
Percent overdiagnosis (first screen)	7.8	0.3–26.5
Percent overdiagnosis (second screen)	0.5	0.02–2.1

21,307 women having their second screen, there were 123 screen detected cancers and 58 interval cancers.

Table 2 shows the results of estimation of overdiagnosis (all tumors, invasive and in situ), taking into account the underlying incidence of preclinical disease, sojourn time, and sensitivity. The underlying incidence of preclinical disease was estimated as 3.8 per 1000, which is consistent with the high incidence of breast cancer in the Copenhagen region. The estimated rate of progression to clinical disease in those tumors capable of progression was 0.38, corresponding to a mean sojourn time of about 32 months, similar to estimates in the past (6,7). The sensitivity estimate was at its boundary of 100%. The annual incidence of "overdiagnosable" tumors was estimated as 0.0142 per 1000 (95% confidence interval [CI] 0.0005–0.0493). This corresponds to 7.8% (95% CI 0.3–26.5%) overdiagnosis at the first screen and 0.5% (95% CI 0.02–2.1%) at the second screen, for a total of 30 cancers overdiagnosed in the two rounds. This is 4.8% of the total of 627 cancers diagnosed in the period of observation. When the analysis was restricted to invasive cancers only, very similar estimates of overdiagnosis were obtained: 7.3% at the first screen and 0.5% at the second.

Although similar estimates have been observed in the past in this age group (6), the very small CI on the sensitivity and the very large CI, relative to its magnitude, of the overdiagnosis incidence, suggest that estimation may be unstable, possibly due to overparameterization. We therefore performed three additional analyses, with sensitivity constrained to be 80%, 90%, and 100%, respectively. The results are shown in Table 3. The different sensitivity assumptions had a strong effect on the estimated sojourn time, which increased with decreasing sensitivity. The overdiagnosis estimates were, however, similar to those in Table 2, still with very broad CIs.

DISCUSSION

We estimated the overdiagnosis in mammography screening from the outcome of the first two invitation rounds of the organized screening program in Copenhagen, Denmark. Considering both invasive cancer and carcinoma in situ, we estimated the overdiagnosis at the first screen to be 7.8% with a 95% CI of 0.3–26.5%. At the second screen, the overdiagnosis was estimated to be 0.5% with a 95% CI of 0.02–2.1%. The incidence of preclinical breast cancer was estimated to be 3.8 per 1000 person-years, the mean sojourn time 2.7 years, and the sensitivity close to 100%. Before arriving at any conclusions, we must consider the adequacy of the model applied and the data used.

Adequacy of the Model

The model assumes homogeneity across age groups and over time for all parameters. When we attempted to estimate age-specific parameters, we had problems with convergence, probably due to the relatively sparse data for a complex model with a substantial number of parameters.

Breast cancer incidence is known to increase with age. In 1990, before screening started, the incidence of invasive breast cancers in Copenhagen was 240 per 100,000 for women age 50–59 years and 287 per 100,000 for women age 60–69 years. Mean sojourn time and test sensitivity have been found to increase with age (8), with the main differences seen between women age 40–49 years and women age 50 years or older. Within the age range 50–69 years, variation in sojourn time and sensitivity is modest (6). Splitting the model into separate age groups would be desirable if we had sufficient data, but we believe that the results here are useful as a first evaluation of the parameters of interest.

Table 3. Estimates from Overdiagnosis Modeling with Sensitivity Constrained to 80%, 90%, and 100%

Quantity	Estimate	95% CI
Sensitivity = 80%		
Incidence (true cases)/1000	3.3	2.8–3.7
Mean sojourn time (years)	3.8	3.1–4.4
Incidence of overdiagnosed cases/1000	0.0156	0.0006–0.0528
Percent overdiagnosis (first screen)	8.6	0.3–28.4
Percent overdiagnosis (second screen)	0.6	0.02–2.3
Sensitivity = 90%		
Incidence (true cases)/1000	3.5	3.1–4.0
Mean sojourn time (years)	3.1	2.5–5.7
Incidence of overdiagnosed cases/1000	0.0151	0.0005–0.0548
Percent overdiagnosis (first screen)	8.3	0.3–29.1
Percent overdiagnosis (second screen)	0.6	0.02–2.4
Sensitivity = 100%		
Incidence (true cases)/1000	3.8	3.3–4.2
Mean sojourn time (years)	2.6	2.2–3.1
Incidence of overdiagnosed cases/1000	0.0141	0.0005–0.0520
Percent overdiagnosis (first screen)	7.8	0.3–27.5
Percent overdiagnosis (second screen)	0.5	0.01–2.2

Breast cancer incidence has been increasing in Denmark since the 1960s, but this background trend will have a limited effect in the present analysis, which uses data for only a 6 year period. An individual with two screens is followed up for only 4 years. The age distribution at the second screen will be slightly different from that of the first, as the first screen includes women age 50–71 years at screening, whereas the second screen includes women age 52–71 years at screening. The effect of this on the estimates is marginal.

The sensitivity could be higher in the second screen, where first-screen mammograms were available for comparison. In turn, this would affect the estimation of overdiagnosis. Underestimating the sensitivity of the first screen would lead to overestimating the degree of overdiagnosis, and vice versa. The similar estimates for different constrained values of sensitivity suggest that this is not a problem.

Sojourn time was assumed to be exponentially distributed. Day and Walter (3) found this distribution gave a reasonable fit to breast cancer data. For overdiagnosed tumors, we assumed exponential time to incidence and a sojourn time extending beyond the lifetime of the host. These may be incorrect assumptions, although alternative overdiagnosis models give similar estimated proportions of overdiagnosis. The assumptions remain a qualifying factor, and interpretation must be cautious.

Adequacy of the Data and Compatibility with Other Studies

Four data points were used for the analysis: the number of screen-detected breast cancers at first and second screens, and the number of interval cancers after the first and second screens. This may not be sufficient to identify all four parameters. However, the secondary analyses with sensitivity constrained to 80%, 90%, and 100% showed that the overdiagnosis estimates are reasonably robust.

The estimates depend on the thresholds of malignancy at screening and clinical diagnosis as well as on age distribution, time period, and risk factors. Because of these factors, the estimates are program and population specific. It is therefore important to base the estimation on the data of the specific program. The disadvantage is that estimates are not necessarily valid for other programs and populations than the ones studied. The screening technology, however, is the standard for the times, and the screening regime is similar to others used in developed countries. One might therefore expect our estimates to be in the same range as those from other screening programs.

Our estimates of mean sojourn time are between 2.6 years and 3.8 years, which is typical for this age group (6–8). In the unconstrained analyses, sensitivity was estimated to be close to 100%, which seems high, although again, there is a precedent for this in this age group (6).

The technical point underlying the high estimates of sensitivity is that the component of likelihood relating to the prevalence screen is monotonic increasing in the sensitivity parameter. The likelihood components for the incidence screens are not necessarily monotonic increasing in sensitivity in principle, but are so in practice for many screening datasets. The interval cancer likelihood components are definitely not monotonic increasing in sensitivity, but if there is a relatively small number of interval cancers, the likelihood components relating to the screen-detected cancers predominate, pulling the sensitivity estimate toward its boundary of 100%. Possible alternative approaches to this problem include estimation of sensitivity from the interval cancer data only (8) or assuming a distribution of incidence of "true" cancers which gives a smaller rate of de novo preclinical tumors very soon after a negative screening test. The development and testing of these approaches is a focus of ongoing research.

Our estimated preclinical incidence is high, at 3.8 per 1000 per year. This is consistent with the high incidence of breast cancer in the Copenhagen area (9). Estimates of overdiagnosis in the literature have a wide range (10). Our estimate of around 7–8% at first screen and less than 1% at subsequent screens is consistent with the modest estimates observed in studies which, like ours, take into account the lead time of the tumors and identify screened cohorts (6,11,12). However, the CIs on our estimates of overdiagnosis are wide. This quantity is difficult to estimate with precision and it is likely that a meta-analytic approach using large quantities of screening data from different programs will be necessary to estimate overdiagnosis with high precision.

We observed a slight reduction in our overdiagnosis estimates with exclusion of in situ cases, from 7.8% to 7.3% at first screen. This suggests that 10% of in situ cases are overdiagnosed at first screen, compared to 7.3% of invasive cases. The relatively small difference between the two is surprising, as others have found overdiagnosis to be particularly prevalent in in situ disease (11,12). Our results are, however, consistent with those of McCann et al. (13), who found a deficit in incidence of invasive cancers after the upper age range limit for screening, which was more fully accounted for by earlier detection of in situ cancers as well as invasive. McCann et al. (13) concluded from this that most of the in situ cases would have progressed to invasive disease if they had been left untreated.

As stated in the introduction, overdiagnosis is an extreme form of length bias. A possible approach for future

methodologic development would be a continuous frailty parameter attached to the sojourn time.

CONCLUSION

This article demonstrates the feasibility of estimation of overdiagnosis, taking into account sojourn time and screening sensitivity. It provides evidence that sojourn time and sensitivity in the Copenhagen program are comparable to those observed in other mammography screening programs. The analysis indicated some overdiagnosis of 7–8% at the first screen, and very little overdiagnosis of 0.5% at the second screen. But considerable uncertainty remains as the CIs are very broad. Larger, possibly meta-analytic studies are required to derive precise estimates of overdiagnosis in mammography screening programs.

Acknowledgments

The research reported in this article was undertaken during the tenure of a Research Training Fellowship awarded to Anne Helene Olsen, PhD by the International Agency for Research on Cancer. The study was also funded by the Danish Cancer Society.

REFERENCES

1. Svendsen AL, Olsen AH, von Euler-Chelpin M, Lynge E. Breast cancer incidence after the introduction of mammography screening: what should be expected? *Cancer* 2006;106:1883–90.

2. Olsen AH, Njor SH, Vejborg I. *et al.* Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study. *BMJ* 2005;330:220–22.

3. Day NE, Walter SD. Simplified models of screening for chronic disease: Estimation procedures from mass screening programmes. *Biometrics* 1984;40:1–13.

4. Spiegelhalter D, Thomas A, Best N, *et al.* WinBUGS 1.4.1. Cambridge: MRC Biostatistics Unit, 2004.

5. Duffy SW, Tabar L, Vitak B, *et al.* The relative contributions of screen-detected in situ and invasive carcinomas in reducing mortality from the disease. *Eur J Cancer* 2003;39:1755–60.

6. Chen HH, Duffy SW. A Markov chain method to estimate the tumour progression rate from preclinical to clinical phase, sensitivity and positive predictive value for mammography in breast cancer screening. *Statistician* 1996;45:307–17.

7. Paci E, Duffy SW, Giorgi D, *et al.* Population-based breast cancer screening programmes: estimates of sensitivity, over-diagnosis and early prediction of the benefit. In: Duffy SW, Hill C, Esteve J, eds. *Quantitative Methods for the Evaluation of Cancer Screening.* London: Arnold, 2001:127–35.

8. 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; 20:852–58.

9. Andreasen AH, Andersen KW, Madsen M, Mouridsen H, Olesen KP, Lynge E. Regional trends in breast cancer incidence and mortality in Denmark prior to mammographic screening. *Br J Cancer* 1994;70:133–37.

10. Møller B, Weedon-Fekjær H, Hakulinen T, *et al.* The influence of mammographic screening on national trends in breast cancer incidence. *Eur J Cancer Prev* 2005;14:117–28.

11. Paci E, Warwick J, Falini P, Duffy SW. Overdiagnosis in screening: is the increase in breast cancer incidence a cause for concern? *J Med Screen* 2004;11:23–27.

12. Yen M-F, Tabar L, Vitak B, Smith RA, Chen HH, Duffy SW. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer* 2003;39:1746–54.

13. McCann J, Treasure P, Duffy S. Modelling the impact of detecting and treating carcinoma in situ in a breast screening programme. *J Med Screen* 2004;11:117–25.