

ORIGINAL ARTICLE

Modelling the impact of changes in sensitivity on the outcomes of the UK breast screening programme

Paul Taylor

J Med Screen 2010;17:31–36

DOI: 10.1258/jms.2009.009048

See end of article for
author's affiliations

Correspondence to:
Paul Taylor, Centre for
Health Informatics and
Multiprofessional
Education, University
College London, Highgate
Hill, London N19 5LW, UK;
p.taylor@chime.ucl.ac.uk

Accepted for publication
15 December 2009

Objectives To assess the impact on breast cancer mortality of improving the sensitivity of breast screening programmes.

Methods A Markov model was populated with data obtained from published statistics describing the UK National Health Service Breast Screening Programme and the incidence and mortality of breast cancer in the UK. The model was used to study the impact of changes to the sensitivity of screening. The effects on cancer detection rates and breast cancers and total mortality was studied for a cohort of women followed from age 45 to age 89.

Results Running the model from age 45 to 89, with sensitivity set at the baseline value of 75%, predicts the detection at screening of 44 cancers per thousand of population and the detection outside screening of 82 cancers per thousand of population. Running the model with values of sensitivity from 75% to 95% shows the proportion of cancers detected at screening increasing as screening improves, and deaths from breast cancer falling. The drop in breast cancer deaths is however modest. Increasing sensitivity from 75% to 85% reduces the number of breast cancer deaths from 28 to 27 per thousand.

Conclusions Likely achievable improvements in the sensitivity of screening do not have a marked effect on breast cancer mortality.

INTRODUCTION

Computer-aided detection (CAD) systems analyse digital mammograms and alert the user to regions of the image more likely to contain an abnormality. One question that is often asked of CAD is whether it allows a single reader to achieve the sensitivity normally attained by double reading mammograms. Attempts to compare CAD with double reading have largely failed to achieve either methodological rigour or statistical power with the notable exception of the CADET II trial,^{1,2} which suggested that single reading with CAD provides sensitivity comparable with double reading.

To assess the importance of sensitivity in screening, we need to understand the relationship between sensitivity and the potential impact on outcomes. Clearly, better screening should have a greater impact on mortality, but it is naïve to assume that each extra cancer detected at screening is a 'life saved'.

Questions that are difficult to answer empirically can be addressed using mathematical simulations. Such modelling exercises have been carried out to assess the financial value of digital mammography and of CAD.^{3,4} This paper presents a mathematical modelling exercise, which takes values from official UK statistics and the academic literature

and simulates the impact of changing the sensitivity of screening.

METHODS

Model

The development of cancer in a screened population is modelled as a Markov process. The states and the allowed transitions are shown in Figure 1. Each transition is associated with a probability. Using this model, a cohort of women is followed from age 45 to 89. Each year women are moved from state to state according to the transition probabilities. Examples of age-related transition probabilities for women aged 45, 60 and 75 are shown in Table 1. The transition probabilities can be adjusted to model the impact of different scenarios on outcomes of interest: numbers of screen-detected versus symptomatically presenting breast cancers, non-screen-detected breast cancers, numbers of breast cancer deaths and total deaths. The rest of this subsection provides a short explanation of the model. The next describes how the transition probabilities were determined.

Between the ages of 50 and 64, a proportion of the cohort attends for screening every three years. Two categories of

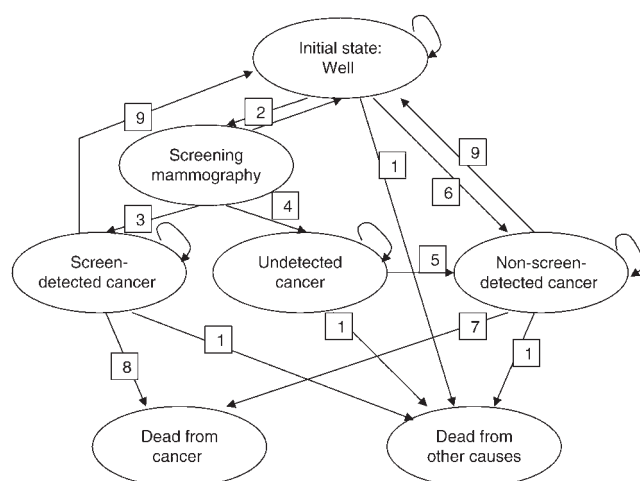


Figure 1 Markov model of breast cancer screening. Ovals represent states through which women pass. Arrows show transitions between states. The numbers on the arrows refer to the numbered paragraphs in the section Model data

detected cancer are considered: detected at screening and detected outside screening. No distinction is made between cancers detected in the intervals between screening visits and those detected in women who have not attended for screening. The progression of disease is not modelled explicitly, but detected cancers are assigned a prognostic category, according to a profile of prognostic categories for the method of detection. The difference in the profiles of screen-detected and non-screen-detected cancers is the basis, in the model, for the benefit due to screening: women with cancers detected at screening are more likely to survive because these cancers are more likely to be assigned to a category with a better prognosis. All cancer deaths are assumed to occur within 10 years of diagnosis, after which patients are returned to the initial state.

If the model is to be used to assess improvements to the screening programme, it must include a pool of detectable but currently undetected cancers. The model therefore includes an 'undetected cancer' state to which women are transferred with a probability reflecting the sensitivity of the screening programme. These cancers are undetected until the 'sojourn' time elapses, after which they present clinically and the woman transfers to the 'non-screen-detected cancer' state. Estimates of sensitivity vary; however, the accuracy of the baseline estimate of sensitivity used is not crucial for the model (Figure 1). If an estimate is chosen for [4] and used to calculate a figure for [5], then so

long as the figure for [6] is set so that the total for [5] and [6] is accurate, the accuracy of the estimate at [4] is of no consequence. Adjustments in the accuracy of screening can now be modelled by varying the proportion of cancers that are undetected.

Model data

This section explains how data were obtained to model the existing screening programme. Authoritative estimates of transitions [5], [7] and [8] could not be obtained. In the case of [5] a variety of estimates were pooled and the sensitivity of the model to the accuracy of the resulting estimate considered, while in the case of [7] and [8] two alternative approaches to estimating the probabilities were compared. Details of these are given below and the results of the comparison are given at the end of this section.

[1] UK mortality statistics give an age- and gender-specific figure for all-cause mortality and a figure for breast cancer mortality in five-year age bands (total deaths coded as C50, breast neoplasm, or D05.1, intraductal carcinoma *in situ*).⁵ Subtracting the breast cancer deaths from the all-cause mortality and dividing by the resident female population in that age range gives the rate at which women die of 'other causes'.

[2] Figures from the UK National Health Service Breast Screening Programme (NHSBSP) for screening uptake in each age group are used to determine the number of attendees in the cohort each year, on the assumption of a three-year screening interval.⁶ Screening is focused on the 50–64 age group, but small numbers are screened in the age groups 45–49, 64–70 and 70 and over.

[3] The proportion of women with cancers detected at screening, for each age group is taken from data published by the NHSBSP.⁶

[4] A proportion of cancers are considered to be 'missed' at screening. This is initially set at a sensitivity of 75%, in line with estimates.⁷

[5] Women stay in the missed cancer state – unless they die of 'other causes' – until the sojourn time has elapsed. An estimate of mean sojourn time (MST) at each age is used to generate a Poisson distribution from which the number of missed cancers presenting each year is determined. Missed cancers are attenuated with the force of the other cause mortality rate (transition [1]) each year that they stay in the missed cancer state. In order to determine age-related estimates of MST, PubMed was searched using relevant keywords and five papers containing seven sets of estimates for the age-related MST obtained.^{8–12} These contained a total of 22 individual estimates of MST for an age band. Plotting age against MST and fitting a simple exponential curve made it possible to generate a figure for MST at each age. These ranged from three years at age 50 to six years at age 75. To test the sensitivity of the model to this estimate, the performance of the model was assessed as the fitted curve was varied through plus and minus 40%. Details of how performance was assessed are given below.

[6] Subtracting the figure in [3] for screen-detected cancer from the total figure for new cancer cases gives a total figure for non-screen-detected cancer for each age

Table 1 Age-related transition probabilities for women aged 45, 60 and 75

Age	Other cause mortality rate, per 100,000	Screen-detected cancer rate, per 100,000 of population	Non-screen-detected cancer rate, per 100,000 of population	Mean sojourn time in years
45	124	6	164	3
60	503	207	96	4
75	2507	65	248	6

group.¹³ Subtracting from this total the number of women presenting via transition [5] in each year gives a figure for presenting cancers not detectable at screening.

[7 and 8] Profiles for screen-detected and non-screen-detected cancers are used to assign each cancer a 'prognostic category'. Each category is associated with a separate survival curve. The modelled impact of screening on mortality will depend very closely on these profiles and associated survival curves. If the profile of cancers is similar for screen-detected and non-screen-detected cancers, screening can have little benefit. If the survival curves for the different prognostic categories are not widely separated, screening can have little benefit. Two approaches were compared: one based on the staging of cancers and one on categories of the Nottingham Prognostic Index (NPI). The prognostic profiles and survival curves for the NPI data are shown in Figure 2. Profiles for stages are based on data from the Malmö screening programme and survival curves were provided by the West Midlands Cancer Intelligence Unit.^{14,15} The NPI profiles were determined from data published by Wishart *et al.*¹⁶ and survival curves based on data published by Blamey *et al.*¹⁷ The latter set of curves is based on a relatively limited number of patients and they were therefore smoothed with a moving average of three. Both profiles were for invasive disease only; estimates of the proportion of detected cancers that are non-invasive in screening and

outside screening were obtained from the NHSBSP and the literature.^{6,18}

[9] An assumption that survival curves plateau at 10 years after diagnosis is enforced, and surviving patients return to the well state.

Most of the above age-related statistics (e.g. cancer registrations) are published for five-year age bands. Others were binned to create five-year age bands that were then used to calculate the various transition probabilities. Simple interpolation was then used to calculate annual transition probabilities, on the assumption that the estimate for the age band is accurate at the centre of the age band and varies smoothly between these points.

As described above, most of the transition probabilities were obtained from UK population statistics. However, the MST and the prognostic profiles (with associated survival curves) were not. In order to compare the performance of the model with different estimates of these values, two 'figures of merit' were used.

First, the model was used to generate an estimate of the overall impact of the screening programme on the number of breast cancer deaths. This was achieved by running a simulation with sensitivity of 0% (transition [4]), simulating the complete absence of screening. Comparing this with a simulation with the baseline sensitivity of 75% allows an estimate of the overall impact of screening on mortality. Two quite different estimates from the literature were used to assess the plausibility of the resulting estimate: the widely accepted estimate from the Swedish two counties trial (31% reduction in breast cancer deaths), and a much smaller estimate from Blanks *et al.*¹⁹ analysis of mortality rates following the introduction of screening (6% reduction in breast deaths).²⁰

Second, age-specific rates of breast cancer mortality were compared with those in the UK mortality statistics and the mean-squared error calculated. Note that the two data-sets should be similar but there is no *a priori* reason why they should be identical, the model simulates a cohort aged 45 in 2006 and followed for 45 years, UK mortality statistics are calculated for the women aged between 45 and 90 in 2006.

Two implementations of the model, one using the profiles and survival curves published for categories of the NPI, and the other based on stages were compared on these two 'figures of merit'. The NPI model gives an estimate for the impact of NHSBSP of a 16% reduction in breast cancer deaths. The stage model gives an estimate of 9%. This latter figure is slightly above the low estimate of Blanks *et al.*¹⁹ The mean-squared error when actual values of age-specific mortality rates are compared with predicted values was 22% using the NPI model compared with 17% using the stage model. Both sets of data seem consistently to underestimate the number of deaths compared with the 2006 data, and to markedly underestimate deaths in older women. On balance the NPI data seemed more robust, with a more plausible estimate of the impact of screening and was used in the remaining simulations. The NPI data were, however, calibrated with the figures for breast cancer deaths in 2006, in order to allow for a possible interaction between age and mortality, resulting in a mean-squared error of 1.1% in the central region. The resulting data are shown in Figure 3.

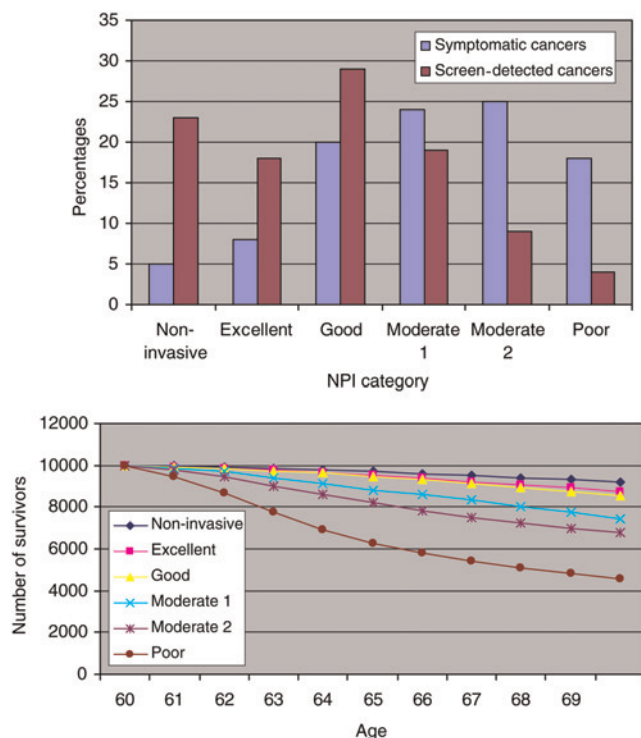


Figure 2 Prognostic profiles for screen and non-screen-detected cancers. The bar chart shows the percentages of non-screen-detected and screen-detected cancers in the different prognostic categories of the Nottingham Prognostic Index (NPI). Cancers in the simulation are assigned to a prognostic category according to these frequencies, depending on the method of detection. The graph shows the survival curves, for an imaginary cohort of 10,000 women aged 60 in 2006 in each of these categories. The curves are based on data for cancer deaths collated by Nottingham City Hospital combined with 'other cause mortality' rates, for ages 60 to 69, calculated from UK mortality statistics

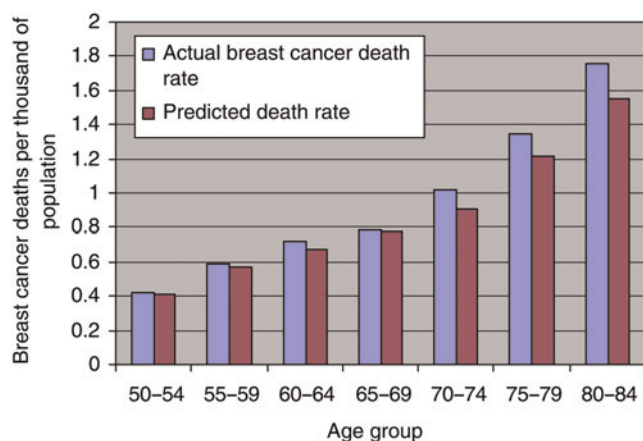


Figure 3 Comparison of (a) actual breast cancer deaths for women age 50–84 in 2006 and (b) predicted breast cancer deaths using the Nottingham Prognostic Index model for the population aged 45 in 2006

The NPI model was run varying the estimates of MST through plus or minus 40% around the original estimate. The estimated impact of screening varied from 18% at –40% (MSTs of 2–3 years) to 16% at +40% (MSTs of 5–8 years). The model therefore seems robust to likely errors in the estimate of MST.

The model was used to simulate changes in the cancer detection rate that might follow from changes in the screening protocol. The model was run varying the sensitivity from 75% to 95%. Each simulation was run 1,000,000 times.

RESULTS

Running the model from age 45 to 89, with sensitivity set at the baseline value of 75%, predicts the detection at screening of 44 cancers per thousand of population and the detection outside of screening of 82 cancers per thousand of population. Improving screening to a sensitivity of 85% should increase the number of cancers detected at screening to 50 per thousand and decrease the number detected outside screening to 76 per thousand. However, the number of breast cancer deaths moves only slightly, from 28 to 27 per thousand. Table 2 shows the impact of changing the sensitivity of screening over a wider range, from 75% to 95%. The change in cancer detection rates is clear and there

is a modest drop in breast cancer deaths as screening improves over this range, but almost no impact on overall survival is observed.

DISCUSSION

The original aim of this work was to assess the possible impact of CAD on the outcomes of the screening programme. In order to determine how CAD might alter the sensitivity of screening, a pair of linked systematic reviews was carried out, looking first at comparisons of single reading and CAD and then at comparisons of single and double reading. This work has been described in detail elsewhere.²¹ Ten studies were found comparing single reading with single reading with CAD, and 17 comparing single reading with double reading. Double reading with arbitration (films where the two initial readers disagree are reviewed by a third) shows a significant increase in detection rate (odds ratio: 1.08, 95% CI 1.02–1.15) and a significant decrease in recall rate (odds ratio: 0.94, 95% CI 0.92–0.96). CAD studies do not show a significant increase in cancer detection rate (odds ratio: 1.04, 95% CI 0.95–1.13) and – although there is considerable heterogeneity between studies – show an increased recall rate (odds ratio: 1.10, 95% CI 1.09–1.12). These differences in cancer detection rate are relatively small (expressed as a risk difference the best estimate of the effect due to CAD is 0.16 extra cancers per 1000 women screened, for double reading 0.44 extra cancers per 1000 women screened). Rather than attempt to compare the model run on such similar estimates, it seemed more illuminating to use the model to simulate larger changes in sensitivity. The conclusion is that the kinds of improvements in sensitivity that we can anticipate are unlikely to have a significant impact on the main outcomes of interest to the screening programme.

Perhaps the best evidence about the impact of CAD is from the CADET II trial. This was designed as an equivalence trial and powered to detect a 10% difference between the intervention and control conditions.² No difference was detected so the two are considered equivalent. The systematic review of other trials suggests that neither CAD nor double reading is likely to improve screening sensitivity by 10%. The modelling reported here suggests that improvements of less than 10% are not going to have an impact on the chief outcome of interest to screening programmes.

One conclusion to draw from this is that proposed enhancements of screening should target the specificity of

Table 2 Outcomes per 100,000 of population for different levels of sensitivity of screening

Sensitivity of screening	Screen-detected cancers	Non-screen-detected cancers	Breast cancer deaths	Total deaths
95	5534	7143	2647	71,317
92.5	5389	7260	2682	71,212
90	5255	7374	2696	71,291
87.5	5096	7521	2690	71,260
85	4953	7642	2728	71,333
82.5	4816	7785	2748	71,388
80	4694	7879	2751	71,308
77.5	4505	8063	2798	71,338
75	4370	8246	2816	71,353

screening rather than the sensitivity, where the scope for real improvement is perhaps limited. By itself, this would argue for the retention of double reading and not single reading with CAD. Note that both the systematic review and the CADET II trial showed a significant increase in recall rate with CAD, although in the case of the CADET II study this was due to an effect at only one of the three centres studied.² CAD can be justified on cost grounds if the saving in radiologist time of not doing double reading is less than the combined cost of CAD and of dealing with the additional recalls generated by CAD. These additional recalls must also be considered as a decrement in quality from a patient's perspective.

The different versions of the model tested suggested that screening reduced breast cancer deaths by between 9% and 17% in the cohort as whole. The percentage reduction among women attending all screening visits will be higher. The range of estimates is lower than the widely quoted figure of 31% from the Swedish two county study and higher than the estimate obtained by Blanks *et al.*¹⁹ from an analysis of changes in mortality rates immediately following the introduction of screening.²⁰ The basis for the impact of screening on mortality is shown graphically in Figure 3: screen-detected cancers are more likely to be in categories that have better survival curves. Blamey *et al.*¹⁷ have demonstrated improvements in survival of patients in all prognostic groups since the 1980s. Such improvements bring the curves closer together, which must mean that the scope for reducing mortality is less than it was in the 1980s, when the Swedish trial started. It should also be noted that the potential benefit shown by the separation of the curves will be attenuated as a consequence of 'other cause' mortality, and this attenuation is greater in older women. On the other hand, as Blanks *et al.* note, significant technical and procedural improvements in screening should mean that the programme is operating more effectively than when it was introduced, increasing the difference between the profiles of screen-detected and non-screen-detected cancers.

A better estimate of the impact of screening on mortality seems to be obtained using profiles and survival curves for the NPI rather than for cancer stage. This might well reflect that the NPI is a better approach to classification of prognosis than stage. It should, however, be borne in mind that the only estimate found of the distribution of stages across screen-detected and non-screen-detected cancers was taken from a relatively small data-set (324 invasive screen-detected, 617 invasive non-screen-detected cancers compared with 2265 non-screen-detected and 1852 screen-detected in the NPI data-set).^{14,16} The model also requires survival curves for prognostic classes. Here the situation is reversed: the West Midland Cancer Intelligence unit provided survival curves passed on data from 12,701 cancers, whereas the NPI curves were calculated on a sample of 2226.^{15,17} It should be noted that both the NPI and the stage data seem to lead to underestimates of mortality when compared with the 2006 mortality statistics, especially in older women. This may reflect an interaction between prognostic category and age.

The model does not perform better at high estimates of sojourn time. This may seem puzzling, since longer sojourn times increase the window of opportunity for detecting a cancer at screening. That effect is not captured

in this model, which ignores the existence of cancers prior to the screening visit. Sojourn time is used solely to measure the duration between a cancer being missed at screening and its subsequent clinical presentation, and hence determines the proportion of cancers that remain undetected because the woman dies of another cause before presentation. Longer sojourn times will tend to lower the impact of screening, as measured by the model, since they allow more missed cancers to disappear in this way. However, the effect seems to be small and the performance of the model is not greatly affected by changing the estimated MST.

CONCLUSIONS

The data generated by the model suggest that screening is beneficial, but that screen-detected cancers do not translate into lives saved as directly as the proponents of screening would wish. As a consequence, improvements in screening have only a small impact on the mortality from breast cancer, the overall mortality and mean length of life.

Author's affiliations

Paul Taylor, Reader in Health Informatics, Centre for Health Informatics and Multiprofessional Education, University College London, Highgate Hill, London N19 5LW, UK

ACKNOWLEDGEMENTS

The study was partly completed as a component of an evaluation of CAD supported by the UK NHSBSP. The help and assistance of Dr Given-Wilson and Dr Potts is gratefully acknowledged.

Declarations: The author has no competing interests to declare.

REFERENCES

- 1 Bennett RL, Blanks RG, Moss SM. Does the accuracy of single reading with CAD compare with that of double reading? *Clin Radiol* 2006;**61**:1023–8
- 2 Gilbert FJ, Astley SM, Gillan MG, *et al.* Single reading with computer-aided detection for screening mammography. *N Engl J Med* 2008;**359**:1675–84
- 3 Lindfors KK, McGahan MC, Rosenquist CJ, Hurlock GS. Computer-aided detection of breast cancer: a cost-effectiveness study. *Radiology* 2006;**239**:710–7
- 4 Tosteson AN, Stout NK, Fryback DG, *et al.* Cost-effectiveness of digital mammography breast cancer screening. *Ann Intern Med* 2008;**148**:1–10
- 5 National Statistics. *Mortality Statistics: Deaths Registered in 2006*. England: Office of National Statistics, 2008
- 6 NHS Breast Screening Programme. *Breast Screening Programme, England: 2005–06*. Leeds: NHS Information Centre, 2007
- 7 Elmore JG, Armstrong K, Lehman CD, Fletcher SW. Screening for breast cancer. *JAMA* 2005;**293**:1245–56
- 8 Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Trelli S. Estimating mean sojourn time and screening sensitivity using questionnaire data on time since previous screening. *J Med Screen* 2008;**15**:83–90
- 9 Mandelblatt J, Schechter C, Lawrence W, Yi B, Cullen J. The SPECTRUM population model of the impact of screening and treatment on US breast cancer trends from 1975 to 2000. *J Natl Cancer Inst Monogr* 2006;**36**:47–55
- 10 Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast* 2006;**12**:338–42

- 11 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
- 12 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
- 13 National Statistics. *Cancer Statistics Registrations: Registrations of Cancer Diagnosed in 2006*. England: Office of National Statistics, 2008
- 14 Zackrisson S, Janzon L, Manjer J, Andersson I. Improved survival rate for women with interval breast cancer. *J Med Screen* 2007;**14**:138–43
- 15 West Midlands Cancer Intelligence Unit. *0–10 year survival for breast cancer by stage, diagnosed in the West Midlands 1990–1994 followed up to the end of 2004, as at December 2008*, private communication (data used in figures published by CRUK in CancerStats – Breast Cancer UK, April 2009)
- 16 Wishart GC, Greenberg DC, Britton PD, *et al.* Screen-detected vs symptomatic breast cancer: is improved survival due to stage migration alone? *Br J Cancer* 2008;**98**:1741–44
- 17 Blamey RW, Ellis IO, Pinder SE, *et al.* Survival of invasive breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990–1999. *Eur J Cancer* 2007;**43**:1548–55
- 18 Evans AJ, Pinder SE, Ellis IO, Wilson AR. Screen detected ductal carcinoma in situ (DCIS): overdiagnosis or an obligate precursor of invasive disease? *J Med Screen* 2001;**8**:149–51
- 19 Blanks RG, Moss SM, McGahan CE, Quinn MJ, Babb PJ. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990–8: comparison of observed with predicted mortality. *BMJ* 2000;**321**:665–9
- 20 Tabar L, Fagerberg CJ, Gad A, *et al.* Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the breast cancer screening working group of the Swedish National Board of Health and Welfare. *Lancet* 1985;**1**:829–32
- 21 Taylor P, Potts HW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer* 2008;**44**:798–807