Risk factors for breast cancer by mode of diagnosis: some results from a breast cancer screening study

JOHN WHITEHEAD AND JACQUELINE COOPER
From the Department of Applied Statistics, University of Reading

ABSTRACT We have investigated factors affecting the probability that a woman with breast cancer participating in a mammographic screening programme will be diagnosed by the screen. Data from a large American case-control study, with subjects drawn from women participating in an annual screening programme, were used. During the screening programme, 409 cases were identified, the mode of diagnosis being screen detection for 331 and interval detection for 78. No significant relationships were found between mode of diagnosis and age, age at menarche, oral contraceptive use, age at first live birth, age at menopause or history of maternal breast cancer. There was a non-significant trend for particular mammographic patterns to be associated with interval detection. However relative risk of breast cancer and probability of interval detection were observed to increase about the time of the menopause. These results suggest that the 3 yearly mammography programme being introduced in the UK might be improved if an extra examination was included around the time of the menopause.

A recurrent issue in breast cancer screening policy is whether there is a subgroup of women for whom screening is especially valuable. Such women might then become the sole recipients of screening, or they might receive more frequent screens than the rest of the population. Most researchers seek to identify women who possess a combination of risk factors which increases their risk of breast cancer, and then consider the effects of restricting screening to this subgroup. When only asymptomatic women are considered, conclusions are usually negative: no appreciable subgroup is at such high risk that differential screening policies would be worthwhile.

A complementary question will be addressed in this paper. Cases alone are considered, drawn from follow up of participants in a breast cancer screening programme. Factors determining the mode of diagnosis (ie, at the screen or during the intervals between screens) are sought. This potentially allows the identification of women for whom the screening programme is proving to be relatively ineffective. Unfortunately, our conclusions are mostly negative and no obvious subgroup is identified. However, we do believe that both the relative risk of cancer and the relative effectiveness of screening need to be taken into consideration in the identification of any subset of women who might receive especially intensive screening.

The probability that a screening examination fails to identify a tumour that is present (the false negative rate) and the time that it takes for a tumour to develop through the stage in which it is detectable only by screening (the sojourn time) both affect the probability that a tumour is detected by screen rather than in the interval between screens. The methods in this paper cannot distinguish between these two situations.

The investigation reported here does throw some light on questions concerning the menopause and breast cancer risk which have been raised by Alexander and Roberts. Their suggestion that tumours develop more quickly during the menopause should be apparent from an increased proportion of interval cancers around this stage of life. This is because the quickly developing tumours will be more likely to progress through the pre-clinical state to clinical manifestation between screens. Evidence for this phenomenon is presented in this paper.

The Seattle Breast Cancer Study

The Seattle Breast Cancer Study was designed to investigate the relationship between a classification of breast type made from the mammogram of an asymptomatic woman and the risk of subsequent breast cancer. The classification was introduced by Wolfe and comprises four distinct types: N1, P1, P2 and DY. The definitions of the four classes concern the pattern of ducts perceptible from the X-ray: N1 breasts exhibit no ductal pattern, P1 breasts a ductal pattern occupying less than one quarter of the breast.
volume, P2 breasts a greater proportion of ductal pattern, and DY breasts exhibit a dense image often obscuring any underlying ductal pattern. The full definitions used in classification during the Seattle Breast Cancer Study are given elsewhere.\textsuperscript{4} Relative risks of breast cancer of 1:20:3:5:3:1 were found for N1: P1: P2: DY and these were statistically significant, but not of the magnitude first claimed by Wolfe.\textsuperscript{5} Further epidemiological results from the study have been reported.\textsuperscript{6}

The subjects investigated in the Seattle Breast Cancer Study were drawn from a cohort of 40,000 women participating in four breast cancer demonstration projects in Ann Arbor, Honolulu, Seattle and Tucson.\textsuperscript{7} Film-screen mammography was used at Ann Arbor and Honolulu and xerography at Seattle and Tucson. Women in the projects were asymptomatic volunteers receiving a course of five consecutive annual mammograms. (During the course of the study mammography was discontinued for women aged less than 50 due to fears concerning risks of exposure to X-rays.) The data used in this report exclude prevalent cases, defined as those diagnosed at the first screen, in order to minimise the effects of length biased sampling.\textsuperscript{8}

Initial mammograms were taken between September 1973 and July 1976. Women were followed up annually, on the anniversary of their initial mammogram, and all cases identified were recruited into the study. Their initial mammograms were then classified without knowledge of disease status and a detailed questionnaire administered. Two controls were selected for each case, from the same clinic, within the same 5 year age group, and with a similar date of initial mammogram, and corresponding data were assembled for them.

**Interval and screen detected cases**

A total of 706 cases were identified in the study, excluding diagnoses made as a result of the initial screen. Of these, 331 were identified at the second or subsequent screen (screen detected) and 78 were detected during the intervals between screens (interval detected). The remaining 297 were diagnosed after the five year screening programme had ended. Screens involved both mammography and physical examination: of the 331 screen detected cases, 114 were detected by the mammogram alone, 140 by both examinations and 77 by physical examination alone. Thus the assessment of screening given here relates to the combined use of mammography and physical examination and not just to mammography itself.

The results in this section concern the probability that a case is interval detected rather than screen detected. Only the 409 cases detected during the screening programme are considered; the post-screening cases are not included. Overall 78 cases (19\%) were interval detected. The relationship between the probability of interval detection and various risk factors is investigated using logistic regression analysis. Values of the risk factors are drawn from the questionnaire completed on entry to the study, when these women were all asymptomatic. Only family history and age at menopause were subsequently updated. Wolfe pattern was not. Differences between probabilities are expressed in terms of odds ratios.

**Table 1 Mode of diagnosis by age at first live birth for 404 breast cancer cases from the Seattle Breast Cancer Study. Data are numbers of cases (%).**

<table>
<thead>
<tr>
<th>Age at first live birth (years)</th>
<th>Nullip</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 22</td>
<td>66</td>
<td>329</td>
</tr>
<tr>
<td>22-24</td>
<td>69</td>
<td>(83)</td>
</tr>
<tr>
<td>25-27</td>
<td>55</td>
<td>(79)</td>
</tr>
<tr>
<td>28-30</td>
<td>40</td>
<td>(89)</td>
</tr>
<tr>
<td>&gt; 30</td>
<td>37</td>
<td>(93)</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>(93)</td>
</tr>
</tbody>
</table>

**Table 2 Mode of diagnosis by oral contraceptive use for 403 breast cancer cases from the Seattle Breast Cancer Study. Data are numbers of cases (%).**

<table>
<thead>
<tr>
<th>Oral Contraceptive use</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen detected</td>
<td>82</td>
<td>220</td>
<td>302</td>
</tr>
<tr>
<td>Interval detected</td>
<td>18</td>
<td>47</td>
<td>65</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>267</td>
<td>367</td>
</tr>
</tbody>
</table>

**Table 3 Mode of diagnosis by Wolfe classification for 409 breast cancer cases from the Seattle Breast Cancer Study. Data are numbers of cases (%).**

<table>
<thead>
<tr>
<th>Wolfe classification</th>
<th>N1</th>
<th>P1</th>
<th>P2</th>
<th>DY</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen detected</td>
<td>33</td>
<td>95</td>
<td>155</td>
<td>48</td>
<td>331</td>
</tr>
<tr>
<td>Interval detected</td>
<td>3</td>
<td>17</td>
<td>44</td>
<td>14</td>
<td>78</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>112</td>
<td>199</td>
<td>62</td>
<td>409</td>
</tr>
</tbody>
</table>
Tables 1 to 3 concern age at first live birth, oral contraceptive use and Wolfe classification. No significant relationship is apparent either between age at first live birth and mode of diagnosis ($\chi^2 = 2.79$, df = 5, NS) or between oral contraceptive use and mode of diagnosis ($\chi^2 = 0.01$, df = 1, NS). For Wolfe pattern the tendency for P2 and DY patterns to be associated with interval detection, noted in Whitehead et al (1985), is once more apparent, the significance level lying between 5 and 10% ($\chi^2 = 6.04$, df = 3, p < 0.1). Notice that the groups at higher risk of breast cancer (P2 and DY) are also the groups for which the screening programme is least effective.

Similar analyses investigated the effects on mode of diagnosis of age, age at menarche, age at menopause, previous personal history of non-breast cancer, breast cancer in mother, height, weight and breast size. None of these displayed a significant relationship with mode of diagnosis. Table 4 shows the relationship between clinic and mode of diagnosis. The probability of interval detection was least at Ann Arbor and Honolulu, with Seattle close behind and Tucson showing a much greater probability. The first two centres were those using film-screen mammography and the latter two used xerography. An analysis was performed comparing these two techniques. The resulting odds ratio of interval detection for xerography relative to film-screen was estimated to be 2.24 (95% confidence interval 1.35, 3.71). This result was not substantially changed by adjustment for the pathology of the cancer.

Table 4  Mode of diagnosis by clinic of 409 breast cancer cases from the Seattle Breast Cancer Study. Data are numbers of cases (%)

<table>
<thead>
<tr>
<th>Screening clinics</th>
<th>Ann Arbor</th>
<th>Honolulu</th>
<th>Seattle</th>
<th>Tucson</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen detected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>96 (86)</td>
<td>97 (87)</td>
<td>83 (64)</td>
<td>55 (63)</td>
<td>331 (81)</td>
<td></td>
</tr>
<tr>
<td>Interval detected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 (14)</td>
<td>15 (13)</td>
<td>16 (16)</td>
<td>32 (37)</td>
<td>78 (19)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>111</td>
<td>112</td>
<td>99</td>
<td>87</td>
<td>409</td>
</tr>
</tbody>
</table>

Menopausal status

Alexander and Roberts (1987) made some interesting suggestions concerning breast cancer risk around the time of the menopause. Their data were drawn from the Edinburgh Breast Screening Trial, and they showed a relative risk of breast cancer for women in the "menopausal state" relative to other women, pre- and postmenopausal together, of 2.98. As they explained, this finding is crucially dependent on the exact definition of "menopausal state": in this case the definition is that the woman's last monthly period occurred between 2 and 12 months previously. They propose a model for the age incidence of breast cancer which includes an increased growth rate for existing tumours at the time of menopause as an explanation for their findings.

The Seattle data are not ideal for investigation of Alexander and Roberts' hypothesis. The precise definition of "menopause" is not explicit, but for each subject the "age at menopause" has been recorded. The age of both the case and her two matched controls at the time of the case's diagnosis can be found, and thus the number of years between menopause and diagnosis computed. The ages are recorded to the nearest year, and so their difference can be inaccurate by as much as one year in either direction. Alexander and Roberts were concerned with the effect of the menopausal state on breast cancer risk. We began by exploring this relationship using conditional logistic regression analysis of the data from both cases and controls. We then investigated the effect of the menopausal state on mode of diagnosis, which appears to be a more direct evaluation of the hypothesis of Alexander and Roberts.

Figure 1 shows estimated relative risks for women classified by years from menopause to diagnosis, for the 1778 women for whom these data are available. The risk is adjusted for age by the matched nature of the data and the figure was prepared using the results of a conditional logistic regression analysis conducted using PROC LOGIST of the statistical computing package SAS. It is apparent that risk decreases with years from menopause to diagnosis, with a sharp rise and fall interrupting the pattern around the time of menopause itself.

Two possible sources of bias in this analysis must be noted. First, women contracting cancer before their menopause are more likely to die and thus be unable to supply a subsequent date of menopause. They will be excluded from the analysis, removing cases disproportionately from the pre-menopausal groups. Without this effect the underlying downward trend in the data would be stronger still. Second, diagnosis and treatment for breast cancer might involve the imposition of an unnatural menopause for some women. This will only affect women for whom the number of years from menopause to diagnosis takes a negative value: their time of subsequent menopause might be altered by treatment but not their menopausal status at time of diagnosis. The only consequence of this on fig 1 would be to alter the values of the first three points while maintaining the same overall relative hazard within the pre-menopausal group.
A comparison of the women in whom one year elapsed from menopause to diagnosis with all others yields a significant relative risk estimate of 1.99 ($\chi^2 = 5.40, df = 1, p < 0.025$). It must be admitted that the particular form of this analysis has been suggested by the data.

Attention is now restricted to cases detected during the 5 years of screening, and to the probability of interval detection. To prepare fig 2, cases have been classified by mode of diagnosis and years from menopause to diagnosis. The proportion of interval detected cases is plotted against years since menopause. The proportion of interval detected cases clearly rises around the time of menopause and falls subsequently. This is in accord with Alexander and Roberts' suggestion of rapid tumour growth during menopause and a return to normal—and therefore a temporary lull in interval detections—thereafter. A comparison of the proportions of interval detected cases in the one year from menopause to diagnosis group and the remainder gives a significance level lying between 5 and 10% ($\chi^2 = 2.91, df = 1, p < 0.10$).

**Implications for screening policy**

In this paper we have pointed out the importance of considering the effectiveness of screening as well as the risk of disease when identifying groups in need of particularly intensive monitoring. Thus, in the case of breast cancer, the joint effects of probability of screen rather than interval detection and relative risk of breast cancer need to be considered.

Consider as an example, the influence of Wolfe pattern on screening policy. Although Wolfe pattern has only a slight effect on the probability of screen detection, this example will serve to illustrate the calculations and considerations which might actually influence policy in the presence of effects of greater magnitude.

The risk of breast cancer for a woman of P2 breast type is estimated to be 3.5 relative to a woman of N1 type. In obvious notation,

$$P(BC | P2) = 3.5 P(BC | N1)$$
Risk factors for breast cancer by mode of diagnosis

However, the probability of screen detection for a $P_2$ woman with cancer under the annual screening policy of the Seattle Breast Cancer Study is 0.78 (Table 3):

$$P(\text{Screen Detection} | \text{BC, } P_2) = 0.78,$$

whereas

$$P(\text{Screen Detection} | \text{BC, } N_1) = 0.92.$$ 

Thus,

$$P(\text{Screen Detection} | P_2) = P(\text{Screen Detection} | P_2, \text{ BC}) P(\text{BC} | P_2)$$

$$= 0.78 P(\text{BC} | P_2)$$

$$= 0.78 \times 3.5 P(\text{BC} | N_1)$$

$$= 2.73 P(\text{BC} | N_1),$$

whereas

$$P(\text{Screen Detection} | N_1) = 0.92 P(\text{BC} | N_1).$$

Thus the risk of a screen diagnosed cancer for a $P_2$ woman is $2.73/0.92 = 2.97$ relative to an $N_1$ woman.

A $P_2$ woman with cancer is disadvantaged relative to an $N_1$ woman with cancer: screening is less likely to detect her tumour. An egalitarian screening policy would seek to monitor $P_2$ women more frequently than $N_1$ women in order to equalise their probabilities of screen detection when cancer is present.

The calculations above show that screening is more likely to detect a tumour in a $P_2$ breast, although the relative risk is not as great as that for incidence of breast cancer itself. An economically optimal screening policy would concentrate on $P_2$ women for this reason. Although in this example both considerations lead to more frequent monitoring for the $P_2$ group, this might not always happen. Changing the relative risk of breast cancer in the example above from 3.5 to 1.15 gives calculations leading to a relative risk of screen diagnosis of 0.975. The egalitarian approach still gives a (slightly) greater frequency of screens to $P_2$ women, but the economic approach now concentrates (to a small extent) on $N_1$ women.

**Conclusions**

When differences in the probability of interval rather
than screen detection are found, there remains a problem in interpretation. The difference might be due to differences in false negative rates. Thus the method of screening is simply less effective for certain subjects. Alternatively, tumour development might be more rapid for the group showing fewer screen diagnoses: the sojourn time of the tumour might be shorter. In either situation more frequent screening will be advantageous, but the benefits of such a scheme can only be predicted if the correct interpretation of the difference is made. Of course, both mechanisms might be present. The analysis methods presented here cannot distinguish between differential false negative rates and differential sojourn times. Methods developed by A Jemain (PhD Thesis, University of Reading) do allow this distinction to be made and allow an appropriate evaluation of the introduction of a differential screening policy to be made.

Figures 1 and 2 lend support to the hypothesis of Alexander and Roberts, who claim that the rate of tumour development is accelerated during the menopause (sojourn time is decreased). An alternative explanation that some change in the breast during menopause increases the false negative rate would also be consistent with the data. Explanations based on menopause being caused by treatment for breast cancer differentially for the interval and screen detected cases are difficult to construct. Breast cancer itself can lead to ovulatory failure, but this is true mainly of late stage cancers.9 Few patients diagnosed during a screening programme are likely to have suffered ovulatory failure before diagnosis. Our findings would indicate that the programme of three yearly mammography being introduced in the UK might be enhanced by the inclusion of an extra examination around the time of menopause itself.

The authors wish to thank the principal investigator of the Seattle Breast Cancer Study, Dr Thomas Carlile, and his co-investigators, Dr F I Gilbert (Honolulu), Dr A J Present (Tucson) and Dr B A Threatt (Ann Arbor) for making the data of the study available. The preparation and revision of this manuscript benefitted from the epidemiological insights and helpful comments of Drs K McPherson, D Mant and T Rohan. During this project, Jacqueline Cooper was supported by an SERC Advanced Studentship, No 86415906.

Address for correspondence and reprints: Dr J Whitehead, University of Reading, Department of Applied Statistics, Whiteknights, PO Box 217, Reading RG6 2AN.

References


Accepted for publication November 1988