PERGAMON

European Journal of Cancer 38 (2002) 568–573

European Journal of Cancer

www.ejconline.com

Assessment of the early impact of the population-based breast cancer screening programme in Florence (Italy) using mortality and surrogate measures

E. Paci[a,*], D. Giorgi[a], S. Bianchi[b], V. Vezzosi[b], M. Zappa[a], E. Crocetti[a], M. Rosselli del Turco[c]

[a]*Unit of Evaluative and Clinical Epidemiology, CSPO, Center for Study and Prevention of Cancer, Scientific Institute of the Tuscany Region, 12 Via di S. Salvi, 50135 Florence, Italy*

[b]*Pathology Deptartment, University of Florence, Florence, Italy*

[c]*Breast Unit, CSPO, Florence, Italy*

Received 9 February 2001; received in revised form 7 September 2001; accepted 17 October 2001

Abstract

The aim of this study was to evaluate the effects by the end of 1999 of the Florence breast screening programme that started in 1990. Approximately 60 000 women (aged 50–69 years) were enrolled from 1990 to 1993. Breast cancer cases diagnosed from 1990 to 1996 were partitioned by the method of detection, classified by their tumour size and nodal status and followed-up for mortality at on the 31 December 1999. Incidence-based mortality in the 50–74-year-old women and advanced carcinomas rates were assessed. Due to low compliance (approximately 60%) and the long enrolment phase, only approximately 35% of the total age-specific population person-years were screened. The number of invasive cases diagnosed was 1122, 17% higher than the 958 expected. After the prevalence screening, a reduction of approximately a quarter in advanced carcinomas was observed in the invited women (Odds Ratio (OR): 0.74; 95% Confidence Interval (CI): 0.55–0.98). In the period 1990–1999, 547 breast cancer deaths were observed: 78 (14%) occurred in women invited and half of these in never responders, 385 (70%) occurred in cases diagnosed before screening started. Disproportionate numbers of deaths occurred in women with advanced tumours. The 19% mortality reduction for the invited women was of borderline statistical significance (observed/expected (O/E) deaths: 0.81; 95% CI: 0.64–1.01); by a one-sided test the result would be unequivocally significant. The mortality reduction attributable to screening in the whole population over the 10-year period was 3.2%. The incidence-based mortality analysis confirmed the current follow-up time is too short for screening to have had a major effect on the breast cancer mortality trends. Screening performance might be improved by a higher level of compliance and shorter interval times, but the estimate of the mortality reduction for the invited and the lower rate of advanced carcinomas confirmed that the effect of the programme is in the expected direction.

Keywords: Screening; Breast cancer; Early diagnosis; Mortality

1. Introduction

Mortality from breast cancer has fallen in several countries in recent years. The most recently available data for women aged 50–69 years in the UK and the USA, confirm a substantial, sharp reduction [1]. Modelling of the falling mortality in England and Wales, where a population-based screening programme started in 1988, estimated that 6.4% of the overall mortality reduction in the population (21%), observed in the period 1990–1998 in women aged 55–69 years, was directly attributable to screening [2]. Reductions to date might be related to the downstaging which occurred in the last 20 years, independent of the organised screening programmes or to improved breast cancer care and treatment.

The evaluation of the mortality reduction due to screening programmes needs a longer follow-up than is yet available for the majority of organised programmes. Consequently, only part of any reduction observed can be attributed directly to organised screening programmes, although an increasing contribution of

* Corresponding author. Tel.: +39-055-63636-24/10; fax: +39-055-679954.

E-mail address: epid1@cspo.it (E. Paci).

0959-8049/02/$ - see front matter
PII: S0959-8049(01)00382-3

screening to the falling rates is to be expected in the future. For this reason, evaluation of the early impact of screening programmes should use other indicators, as well as mortality, to show the effect in the population.

In recent times, controversies about the efficacy of screening have been discussed [3,4]. In our opinion, the reduction of breast cancer mortality has been well established in the eight randomised clinical trials carried out [5] and the challenge today is the evaluation of the impact of screening on the breast cancer incidence pattern (stage) and, in the near future, on mortality rates. In this paper, we present data on incidence and mortality for breast cancer in the city of Florence and we suggest methods for an early evaluation of the impact of the population-based screening programme.

2. Subjects and methods

At the beginning of the 1990s, population-based breast cancer screening programmes started in many European countries. In Italy, screening by mammography was organised regionally and a survey in 1999 estimated organised programmes covered approximately 20% of the Italian target population. In the City of Florence, the invitation of the 50–69-year-old population started in September 1990 and the first screening round was over at the beginning of 1993. Performance indicators have already been reported [6]. The actual interscreening interval was 2.3 years on average (2 years in the protocol) and all women in the target population were invited to attend the subsequent rounds. Women who became 50 years old during the period 1990–1996 or migrated into the area were invited at rounds 2 and 3. At the end of 1996, the third screening round was ongoing. The Tuscany Region Tumour Registry (RTT) registered all breast cancer cases diagnosed in the period 1990–1996; data on characteristics of the tumour and mode of detection were collected. Pathologists carried out a review of all available breast cancer cases. Pathological size was defined as Tis, T1a, T1b, T1c, T2+ and nodal status as negative or positive (N+), according to the Union Internationale Contre le Cancer (UICC)-TNM. The stage was considered as advanced when classified as a UICC stage II+. Breast cancer cases (invasive and non-invasive) were divided into four main categories with sub-classification as follows: (1) Cases diagnosed in the first round of screening (SD1). (2) The 'screening cycle' cases, i.e. the unbiased set [7], including: cases detected at repeated screening tests (SDr); women not attending following the invitation at the first round or having their first test at a subsequent round, who were considered as screen-detected in lapsed responders (SDLr); and breast cancer cases clinically detected after a negative screening test and before the subsequent invitation or the end of the study period, who were considered as interval cancer cases (IC). (3) Cases in non-responders, i.e. clinically detected tumours in invited women who never responded to any invitation up to the end of the study (Nresp). (4) Those cases diagnosed in women not yet invited by 31 December 1992 (NYI). A fifth group included cases in special categories or of uncertain diagnostic modality (Oth).

Population data were available from the Municipality of the City of Florence. Follow-up for mortality continued until 31 December 1999. Breast cancer deaths by year were available from the Regional Mortality Registry, they were linked to the Tumour Registry files and attributed to breast cancer cases diagnosed in the period 1990–1996. They were also partitioned by invitation status (i.e. incident before or after the first round invitation) and by detection modality (incidence-based mortality) [8].

The number of invasive breast cancer cases expected in the invited population in the absence of screening was estimated using the age-specific incidence rates in 1988–1989 and the 1990–1992 rates in women not yet invited. The breast cancer tumour size and nodal status distributions in women not yet invited were considered as the reference for clinically-detected cases. The occurrence of carcinomas *in situ* in the absence of screening was also estimated from the rates in women not yet invited.

2.1. Estimate of the mortality reduction in invited women

Breast cancer-specific case fatality rates were estimated for women not yet invited at each half-year since diagnosis (see the Appendix). Following the classical method from Morrison [9], the estimated case-fatality rate to the end of follow-up was applied to the number of cases expected each year (considering June as the average month of incidence) in the period 1990–1996. For example, the estimated case fatality rate at 9.5 years since diagnosis was applied to the expected cases in 1990. The cumulative number of expected deaths from incident breast cancer cases was compared with the observed deaths (observed/expected (O/E) deaths).

2.2. Early impact of screening

The early impact of screening is attributable to the diagnostic anticipation at the prevalent screening test and to the screening cycle, i.e. the capability of repeated screening of shifting the tumour towards an earlier stage at diagnosis. The number of observed cases was compared in an intention-to-treat evaluation with the expected incidence in the absence of screening according to the formula

$$\text{Observed (IC + SDr + SDLr + Nresp)} = \text{Expected (NYI)} \quad (1)$$

This is based on the approach defined by Day and Duffy as the unbiased set [7], but also includes cases screen-detected in lapsed responders. In theory, the expected occurrence of breast cancer in the invited population just after the detection of breast cancers at the first screening test should nearly equal the number of cases clinically detected in the interscreening interval plus those detected at repeated screening. Cases in women who were lapsed responders were included to complete the yield of cases in the invited-screened cohort. In the intention-to-treat analysis, after the exclusion of the cases detected at the first test (prevalent cases), categories 2 and 3 were grouped and considered as the 'after-prevalence screening set'. The probability of being stage II+ in the 'after-prevalence' screening' set was compared with that expected in women not yet invited using logistic regression. A comparison of the screening cycle series with those expected in the absence of screening was also carried out, given the comparable stage distribution and case fatality rates of the not yet invited and non-responders cases.

3. Results

The average 1990–1996 target population aged 50–69 years was 59 947. Out of the total number of person-years, the estimated number from the time of first invitation until the end of study period (31 December 1996) was 254 890. The proportion of women responding in the first round was 55%, and in the second and third, 60%. Due to the long duration of the enrolment and low compliance, only approximately 35% of the total 1990–1996 population person-years was covered by screening.

3.1. Incidence and tumour characteristics

The expected number of breast cancer cases in the 1990–1996 study period in the target population of the 50–69-year-old resident women in absence of screening was estimated to be 958 invasive carcinomas, while 1122 (17% more) were observed. 84 *in situ* carcinomas were detected.

In Fig. 1, the population-based incidence rates of carcinoma *in situ* (Tis) and invasive T2+ and node-positive tumours are shown for the period 1990–1996. Fluctuations in the number of mammograms performed each year occurred due to organisational reasons (peaks in 1993 and 1995).

In Table 1, the number and proportion of breast cancer cases by method of detection is shown according to the main tumour characteristics. The proportions of carcinoma *in situ* and T1a cases were similar among the three groups of screen-detected cases (Tis+T1a: 21, 26 and 28% in SD1, SDr and SDLr groups, respectively). Among clinically detected cases (IC, Nresp and NYI), the proportions of cases T2+ were 31, 38 and 34%, respectively and node-positive (39, 38 and 38%, respectively) were broadly comparable. 227 breast cancer cases were screen-detected at the first round, of which 55 (24%) were stage II+ carcinomas.

3.2. Early impact of screening

The expected number of cases in the invited population after the prevalence screening was 586 invasive breast cancer cases and 27 Tis. The total number of observed cases in the 'after-prevalence' screening set was 545 invasive and 41 carcinomas *in situ*. The number of advanced carcinomas (stage II+) observed in the 'after-prevalence' screening set was 248 (46% of the invasive carcinomas). Out of the not yet invited invasive cases ($n=307$), 53% were advanced. The logistic odds ratio of being an invasive advanced breast cancer *versus* early cases in the 'after prevalence' screening set, with the not yet invited group as the reference category (age-adjusted), was 0.74 (95% Confidence Interval (CI): 0.55–0.98). In the screening cycle set (excluding the non-responders), 40% of the cases were advanced and the logistic odds ratio was 0.56 (95% CI: 0.41–0.78). The exclusion of the SDLr cases from the screening cycle set did not change significantly the estimates of the proportion of advanced carcinomas (stage II+: 41%). Of the 166 deaths from breast cancer, 128 (77%) occurred in women with advanced tumours at diagnosis. In the not yet-invited women, 86% of the breast cancer deaths occurred in the 53% with advanced tumours. In cases screen-detected in the first round, 62% of the deaths occurred in the 24% with advanced tumours.

3.3. Incidence-based-mortality and estimate of mortality reduction

The observed number of deaths in the whole period 1990–1999 based on breast cancer cases diagnosed in 1990–1996 was 211. Of these, 45 died from causes other than breast cancer and 4 moved to other municipalities before dying or were older than 74 years when dying from breast cancer.

In Table 2, the 1990–1999 breast cancer deaths registered in the Florence City 50–74-year-old female population are presented by diagnostic modality. The cumulative number of deaths was 547. Out of them, 385 (70%) occurred in cases diagnosed before 1990, and 78 (14%) in women diagnosed after the first round screen-

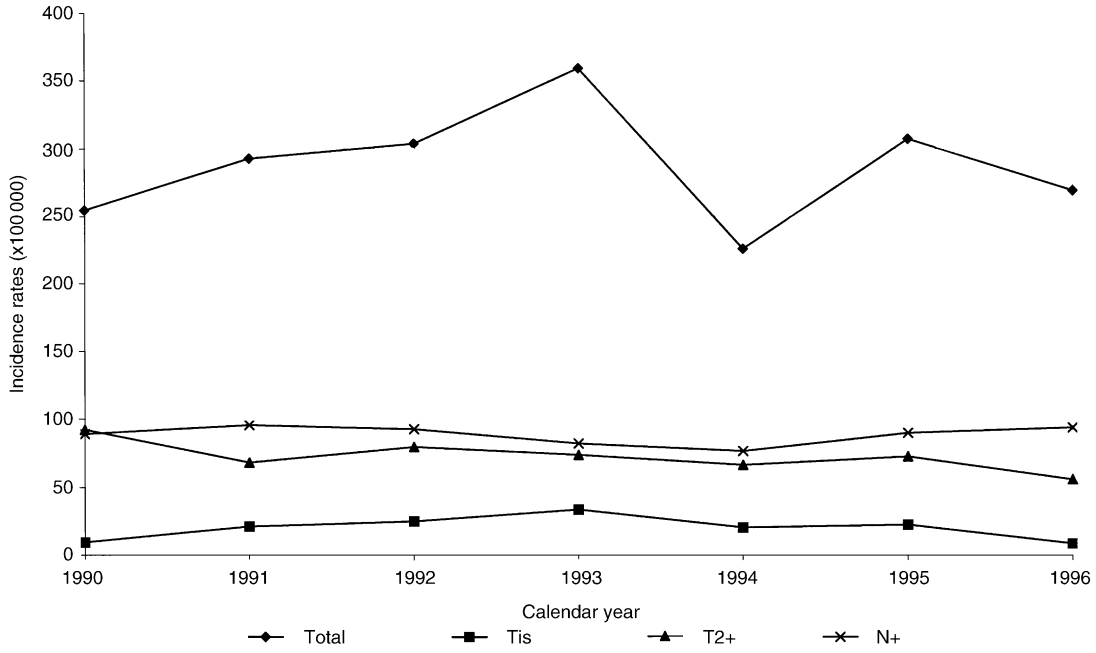

Fig. 1. Incidence of breast cancer cases in the City of Florence (1990–1996) by year and tumour characteristics.

Table 1
Tumour characteristics by method of detection of breast cancer cases in the Florence City Screening Programme (1990–1996)

	Screen-detected at first round	Screen-detected at 2nd and 3rd rounds	Screen detected in lapsed responders	Interval cancers	Never responders	Not yet invited	Other	Total
	n (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)	*n* (%)
Tumour size								
Tis	22 (10)	16 (10)	13 (13)	4 (4)	8 (4)	15 (5)	6 (8)	84 (7)
T1a	26 (11)	27 (16)	15 (15)	8 (7)	7 (3)	14 (4)	4 (6)	101 (8)
T1b	65 (29)	36 (21)	20 (19)	14 (12)	12 (6)	38 (12)	6 (8)	191 (16)
T1c	86 (38)	68 (40)	32 (31)	45 (40)	75 (37)	116 (36)	24 (34)	446 (37)
T2+	24 (11)	20 (12)	19 (18)	35 (31)	76 (38)	110 (34)	21 (30)	305 (25)
Missing or non-operated	4 (2)	1 (1)	4 (4)	7 (6)	24 (12)	29 (9)	10 (14)	79 (7)
Nodal status								
Negative	166 (73)	107 (64)	63 (61)	62 (55)	91 (45)	171 (53)	34 (48)	694 (57)
Positive	41 (18)	40 (24)	22 (21)	44 (39)	77 (38)	122 (38)	25 (35)	371 (31)
Missing or non-operated	20 (9)	21 (13)	18 (17)	7 (6)	34 (17)	29 (9)	12 (17)	141 (12)
Total	227 (100)	168 (100)	103 (100)	113 (100)	202 (100)	322 (100)	71 (100)	1206 (100)

ing invitation. Half of the deaths among the invited ($n = 39$) occurred among the never responders.

The number of breast cancer deaths expected in the population in the absence of screening, based on the breast cancer incidence and case fatality rates in women not yet invited was 180.6 compared with the 162 observed. If we exclude 84 deaths observed in the not yet invited and the 'other' subgroup of breast cancer cases, and a corresponding number among the expected ones, the remaining number of expected deaths in the invited (intention-to-treat analysis) was 96.6 and the O/E ratio: 0.81 (95% CI: 0.64–1.01), which is of borderline statistical significance.

Given an estimated breast cancer mortality reduction for the invited women of 19%, and the 96.6 deaths expected in the absence of screening, the impact of breast cancer screening programme would be 18.4 fewer deaths occurring in the invited population.

If so, the screening programme would have reduced the 50–74-year-old age-specific breast cancer mortality on average by 3.2% ($= 18/(547 + 18)$) over the period 1990–1999.

4. Discussion

In this paper, we confirm that organised breast screening programmes which started around 1990 could have only a little impact within the first 10 years on breast cancer mortality in the whole population of

Table 2
Breast cancer deaths 1990–1999 in women aged 50–74 years and incidence-based-mortality (breast cancer cases detected in the period 1990–1996), by detection modality

Breast cancer deaths	Year of death										Total
	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	*n* (%)
Total	61	50	62	57	59	61	46	56	50	45	547 (100%)
In cases detected in 1990–1996	3	4	10	18	14	27	23	20	25	18	162 (30%)
Not yet invited	1	3	10	14	8	14	7	2	9	3	71 (13%)
Other modalities	2	1	0	0	2	1	0	2	2	3	13 (2%)
Screen-detected at first round	0	0	0	0	0	3	5	3	2	0	13
Screen-detected at the 2nd and 3rd rounds	0	0	0	0	0	1	2	1	1	2	7
Screen-detected in lapse responders	0	0	0	0	0	0	1	1	2	0	4
Interval cancers	0	0	0	2	1	2	2	3	3	2	15
Never responders	0	0	0	2	3	6	6	8	6	8	39
Invited 1st round	0	0	0	4	4	12	16	16	14	12	78 (14%)

women in the screening age groups because of the diluting effect of deaths occurring in women diagnosed before the screening invitation. The evaluation of the mortality reduction due to screening, as is evident from our analysis, needs a longer period of follow-up than is currently available for the majority of organised programmes. Consequently, only a part of any reduction observed so far can be attributed directly to the organised screening programmes, although an increasing contribution of screening to the falling rates is to be expected in the future.

In Florence, the coverage of the screening programme, which started in 1990 was only 35% due to the long enrolment period and low compliance rates, and 70% of breast cancer deaths in 1990–1999 were in cases diagnosed before screening started. Thus, by the end of 1999, screening could have influenced only a small number of the total deaths for breast cancer occurring in the target population; the total impact, however, is likely to be around 7%, i.e. double the average reduction over the period 1990–1999.

The estimated breast cancer mortality reduction for the invited women was 19%, which is in agreement with what would be expected from modelling the impact of the breast cancer screening programme [10,11], and is equivalent to approximately 19 fewer deaths observed in the invited population. This estimate can only partially explain the observed decreasing trend in breast cancer mortality described in the Florence area [12], which occurred largely in breast cancer cases diagnosed before 1990. Reasons for the observed mortality reduction might be found in downstaging or in the new treatments used in the 1980s. Among deaths occurring in breast cancer cases diagnosed after the first invitation in the programme, a large number occurred in the never responding women: the compliance rate in the Florence programme was much lower than has been achieved elsewhere. None the less, the estimated mortality reduction for the invited women (intention-to-treat analysis) suggests that the programme is working in the right direction. If one were sufficiently confident (from previous randomised trial results) that at least screening would not increase mortality, one might perform a one-sided test. In such a case, our reduction in mortality would be unequivocally significant.

Early indicators (i.e. tumour characteristics) have allowed an evaluation of the impact and will offer an opportunity to improve the performance of the programme. The number of cases diagnosed in the 'after prevalence' screening set is comparable to those expected in the absence of screening. We have already shown that on the basis of statistical modelling estimates [13], no excess of breast cancer cases was evident at the prevalent screening. Our estimate suggests that an organised population-based screening programme with moderate levels of detection of carcinomas *in situ* will not result in an important excess of cases due to screening.

In screen-detected cases, the distribution by stage was similar at the first and repeated screening rounds. This result confirms the finding observed in The Netherlands and recently discussed by Boer and colleagues [14]. A possible reason could be the varying interscreening interval—in practice longer than 2 years for organisational reasons. The observed distribution is a possible marker of a lower than expected impact of screening, which in theory, should show a better stage distribution at repeated screening tests. The rates of breast cancer cases by detection modality showed that interval cancers and non-responders had a distribution of tumour characteristics which was comparable to the not yet invited cancer cases.

The proportion of advanced tumours was much lower in the screening cycle set than in women not yet invited, but the high number of cases in the group of never responders strongly determines the whole distribution

of the 'after prevalence' screening set. In the intention-to-treat analysis, the low rates of compliance and the characteristics of non-participants had a strong influence on the overall results.

As appears from this analysis, the impact of breast cancer screening in Florence is largely due to the diagnostic anticipation at the prevalent round and, because of the participation rates and long intervals, the capability of reducing the proportion of advanced tumours over the screening cycle is still lower than is theoretically possible. Better performance of the Florence programme is essential to reduce the proportion of advanced tumours and the mortality rate.

In conclusion, the results of the early evaluation of the programme showed that the Florence programme has not yet had a large impact on overall breast cancer mortality but, as the current follow-up time of the programme is too short, this finding should not be considered as an absence of benefit from the programme. The estimated mortality reduction in the invited/responding women confirmed that the effect of the programme is in the expected direction.

The performance of the screening cycle (stage distribution at repeated tests) and attendance rates set a limit on the global impact of the Florence screening programme. Measures to improve the performance of screening should be planned for the near future.

Acknowledgements

The Florence City Programme was partially supported by the EU screening network contract S12.307923 (2000CVG2-031). Dr Paci was partially supported by a UICC-ICCRET fellowship. We thank N. Susini for data management and I. Poggesi for editing the draft. We also thank a referee for very helpful suggestions.

Appendix

Incidence rates (×1000): 50–54 years: 1.86, 55–59 years: 1.86, 60–64 years: 2.42, 65–69 years: 3.05

Breast cancer-specific survival in women not yet invited:

Years since diagnosis:	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Survival (%):	99	96	93	88	86	83	81	79	77	75

Population and Expected incident cases in the absence of screening

	1990	1991	1992	1993	1994	1995	1996
Population (50–69 years)	62 918	61 854	60 605	59 829	58 879	57 939	57 608
Expected (50–69 years)	145.4	140.6	137.5	135.8	133.4	133.3	132.4

References

1. Peto R, Boreham J, Clarke M, Davies C, Beral V. UK and USA breast cancer deaths down 25% in year 2000 at ages 20–69. *Lancet* 2000, **355**, 1822.
2. 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, 1990–1998: comparison of observed with predicted mortality. *Br Med J* 2000, **321**, 665–669.
3. Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000, **355**, 129–134.
4. Coebergh JWW. Challenges and pitfalls of mass screening in the European Union. *Eur J Cancer* 2000, **36**, 1469–1472.
5. de Koening H. Assessment of nation-wide cancer-screening programmes. *Lancet* 2000, **355**, 80–81.
6. Giorgi D, Ambrogetti D, Bianchi S, *et al*. Design and Preliminary Results of the Florence Breast Cancer Screening Programme (Progetto Firenze Donna). *Eur J Cancer Prev* 1994, **3**(Suppl. 1), 29–34.
7. Day NE, Duffy SW. Trial design based on surrogate endpoints—application to comparison of different breast screening frequencies. *J R Statist Soc A* 1996, **159**, 49–60.
8. Feuer EJ, Merrill RM, Hankey BF. Cancer Surveillance Series: interpreting trends in prostate cancer—part II: cause of death misclassification and the recent rise and fall in prostate cancer mortality. *JNCI* 1999, **91**, 1025–1032.
9. Morrison AS. *Screening in Chronic Disease*. Oxford University Press, New York, 1985.
10. Paci E, Boer R, Zappa M, *et al*. A model-based prediction of the impact on reduction in mortality by a breast cancer screening programme in the city of Florence, Italy. *Eur J Cancer* 1995, **31**, 348–353.
11. van den Akker-van Marle E, de Koning H, Boer R, van der Maas P. Reduction in breast cancer mortality due to the introduction of mass screening in The Netherlands: comparison with the United Kingdom. *J Med Screen* 1999, **6**, 30–34.
12. Barchielli A, Paci E. Recent trends in breast cancer mortality, incidence, survival and mammographic screening activities in Tuscany, Central Italy. *Cancer Causes Control* 2001, **12**, 249–255.
13. Paci E, Duffy SW, Giorgi D, Prevost TC,, Rosselli de Turco M. Population-based breast cancer screening programmes: estimates of sensitivity, overdiagnosis and early prediction of the benefit. In: Quantitive methods for the evaluation of cancer screening. Duffy SW, Hill C, Esteve, J, eds. Arnold, London, 2001.
14. Boer R, de Koning H, van Oortmarssen G, Warmerdam P, van der Maas P. Stage distribution at first and repeated examinations in breast cancer screening. *J Med Screen* 1999, **6**, 132–138.