Stage distribution at first and repeat examinations in breast cancer screening

Rob Boer, Harry de Koning, Gerrit van Oortmarssen, Peter Warmerdam, Paul van der Maas

Abstract

***Objectives*—To investigate observed stage distributions at first and repeat screenings. To compare the observed outcomes with expected values based on simulation modelling, varying the assumptions about the natural history of the disease.**

***Methods*—An overview is made of observed data on stage distribution at first and repeat screenings and the difference between those distributions is summarised in a Gini coefficient. Four possible explanations for the observations are considered, two of these are worked out as Miscan simulation models, and the outcomes are compared with observations.**

***Results*—Often the reported stage distributions at repeat screenings are not or only slightly more favourable than at first screenings and, in the ones that are more favourable, the difference is relatively small. If, in the Miscan model, it is assumed that there is no correlation between the duration of preclinical breast cancer in consecutive tumour size categories and that there is a strong influence of latent cancers, it is not possible to reproduce the observed outcomes.**

***Conclusions*—The two modelled explanations are not sufficient. Decreasing sensitivity seems an unlikely explanation for the discrepancy in many screening programmes. The possibility that the observations may be explained because false reassurance has been given should be seriously considered and investigated.**

(*J Med Screen* 1999;**6**:132–138)

Keywords: breast cancer; theoretical models; computer simulation

Department of Public Health, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, the Netherlands
R Boer, *informatician/statistician*
H de Koning, *associate professor*
G van Oortmarssen, *mathematician*
P Warmerdam, *econometrist*
P van der Maas, *professor*

Correspondence to:
Dr Boer

Accepted for publication
2 September 1999

The performance of cancer screening programmes can be measured by the detection rates—that is, numbers of breast cancers detected per 1000 women screened, and stage distributions in first and subsequent screening rounds and by the incidence and stage distribution of interval cancers compared with a situation without screening. The stage distribution of cancers detected in the first screening round, as a consequence of earlier detection, is expected to be more favourable than that of clinically diagnosed cancers. Because most prevalent cancers are removed from the screened population at the first round, subsequent rounds will only detect cancers that have developed during the screening interval. If the screening interval is sufficiently short, detection rates in subsequent rounds will be much lower than in the first round, and the stage distribution is expected to be more favourable than in the first round. In breast cancer screening programmes, detection rates in subsequent rounds have indeed been found to be much lower than in the first round, but the expected concomitant improvement in stage distribution is often small or even absent.

This article investigates the stage distribution of breast cancers detected at repeat screenings compared with first screenings as reported in the literature, using a simple summary measure for a difference in stage distribution. The observed outcomes are compared with expected values based on simulation modelling. Four possible explanations for the difference between observed and expected values are suggested. The possibilities of further investigation are considered for each.

Methods

Observed data were derived from publications on several breast cancer screening projects,[1–19] summarised in table 1. Some reports split up screen detected cancers by first and repeat screenings, others by first and subsequent rounds, where a subsequent round contains some first screenings of women who were not screened at the first round of invitations.

Different categorisation of size distributions complicates comparison, therefore we summarised the difference in size distribution between first and repeat screening by a single value—the Gini coefficient. This coefficient was originally intended to measure inequality of income in a population.[20] The Gini coefficient G is given by:

$$G = \sum_{i=1}^{N-1} (p_i + p_{i+1})\,(Q_i - P_i)$$

where p_i and q_i denote the proportion of cancers in tumour size category i for first and subsequent screenings respectively,

$$P_i = \sum_{j=1}^{i} p_j \text{ and } Q_i = \sum_{j=1}^{i} q_j$$

are the corresponding cumulative proportions, and N is the number of size categories.

Figure 1 illustrates the calculation of this coefficient for two screening projects. The cumulative distribution of tumour sizes is given for first screenings of cancers on the *x* axis and for subsequent screenings on the *y* axis. The more favourable the size distribution of repeat screenings is relative to the first screening, the higher the graph will be in the figure. The Gini coefficient is the area between the graph and

Table 1 Overview of observed stage distribution by tumour size or UICC stage at first and repeat screenings. Results are shown as percentages

Age range/round	*Screening interval*	*Number of cancers*	*Stage distribution by size or UICC stage (%)*						
Malmo trial[1]									
45–69			Non-inversion	Inversion ⩽10 mm	Inversion >10 mm				
Round 1		118	16	43	41				
Round 2	22 months	58	21	48	31				
Round 3	20 months	46	15	46	39				
Two county study, Kopparberg[2]									
⩾40			Tis	⩽5 mm	⩽10 mm	⩽20 mm	>20 mm		
First round		275	7.6	3.6	28.4	37.5	22.9		
Second round	24/33 months	109	10.1	8.2	20.2	47.7	13.8		
Two county study, Ostergotland[3]									
40–74			Tis	⩽10 mm	⩽20 mm	21–50 mm	>50 mm		
First screening		226	10.2	38.5	34.9	14.2	2.2		
Second screening	24/33 months	106	11.3	36.8	42.5	8.5	0.9		
Two county study[9]									
40–69			1–9 mm	10–14 mm	15–19 mm	20–29 mm	30+ mm		
First screening		284	26.1	29.6	19.7	14.1	10.5		
Second screening	24/33 months	375	27.2	26.4	24.0	16.5	5.9		
Two county study[10]									
40–74			1–9 mm	10–14 mm	15–19 mm	20–29 mm	30+ mm		
First screening		382	26	29	19	15	10		
Later screenings	24/33 months	424	27	27	24	17	6		
Stockholm trial[5]									
40–64			Tis+I	II+					
First round		124	70	30					
Second round	2.3 years	92	77	23					
Edinburgh trial[7]									
45–64			⩽10 mm	11–20 mm	21–50 mm				
Prevalence screening		73	25	47	29				
Incidence screening	1 year (2 years' mammography)	122	29	44	27				
Edinburgh trial[8]									
45–64			pTis	pT1	pT2	pT3			
Prevalence screening		123	19.5	56.1	20.3	4.1			
Incidence screening	1 year (2 years' mammography)	165	12.1	58.2	21.2	8.5			
HIP trial[25]									
40–65			⩽20 mm	>20 mm					
Initial screening		42	40	60					
Repeat rounds	13 months	51	67	33					
Nijmegen pilot project[22]									
35–64			DCIS	⩽9 mm	10–19 mm	⩾20 mm			
Round 1		74	12	16	32	39			
Rounds 2–4	2 years	120	8	25	43	24			
Utrecht pilot project[22]									
50–64			DCIS	⩽9 mm	10–19 mm	⩾20 mm			
Round 1		108	14	25	40	21			
Rounds 2–4	1, 1.5, 2, 4 years	81	6	32	46	16			
Aurich and Braunschweig pilot projects[11]									
35–84			CIS	<10 mm	10–19 mm	20+ mm			
First screening		109	10	27	38	24			
Second screening	1.1 years	25	8	16	32	44			
Uppsala screening programme[12]									
40–69			1–9 mm	10–14 mm	15–19 mm	20–29 mm	30–49 mm	50+ mm	
First round		153	11.1	31.4	24.8	25.5	4.6	2.6	
Second round	Unknown	127	21.3	43.3	22.8	8.7	3.1	0.8	
Dutch screening programme[13]									
50–69			DCIS	T1a	T1b	T1c	T2	T3,4	
First screening		1754	14.7	6.3	21.1	37.9	17.9	2.1	
Second screening	2 years	449	14.4	5.6	23.3	37.8	17.8	1.1	
San Francisco screening practice[14]									
Age distribution unknown			0	I	II	III	IV		
Initial screening		222	30.2	46.8	21.2	0.9	0.9		
Subsequent screening	N/A	111	26.1	56.8	15.3	1.8	0.0		
Finnish screening programme[15]									
50–64			Tis	T1	T2–3				
First round		135	14.1	64.4	21.5				
Second round	2 years	39	7.7	66.7	25.6				
Florence district screening programme[16]									
40–69			Tis	T1	T2	T2+			
First screening		77	9.1	50.6	31.2	9.1			
Second screening	2.5 years	126	6.3	50.1	39.7	3.9			
Navarra screening programme[17]									
50–64			CIS	T1a	T1b	T1c	T2+		
First screening		286	16.6	4.7	22.4	33.2	23.1		
Second screening	2 years	136	11.9	7.5	23.9	31.3	25.4		
North West England screening programme[18]									
50–64			DCIS	Microinvasive	Other T1a	T1b	T1c	T2+	
First screening		392	15.1	2.6	6.6	30.6	30.4	14.8	
Second screening	3 years	92	20.7	4.3	7.6	27.2	28.3	12.0	
Ontario screening programme[19]									
50+			In situ	Microinvasive	<10 mm	10–14 mm	15–19 mm	20–49 mm	50+ mm
Initial screen		1189	13.3	2.6	17.7	23.3	15.8	25.6	1.8
Rescreens	2 years	336	18.8	6.0	20.8	22.0	17.6	14.6	0.3

the diagonal multiplied by two, the area below the diagonal is counted as negative. When both size distributions are equal, the Gini coefficient is 0. Increasingly favourable size distributions of repeat screenings relative to first screenings are reflected in higher Gini coefficients.

Figure 1 shows the diagonal, which represents equal stage distribution, and two extreme cases of a difference between size distribution at first and repeat screenings.

The 95% confidence intervals for the Gini coefficients were estimated by the bootstrap method.[21] This method ignores any dependency of stage distribution at first and repeat screenings, but we expect that the influence of that on the variability estimate is negligible.

The computer simulation package Miscan for breast cancer screening evaluation is used for generating expected values for stage distribution at first and repeat screenings.[22 23] Miscan incorporates the natural history of the disease, the epidemiology, the design of the screening programme, and the performance of screening.[24] The natural history of breast cancer is modelled as a progression through a number of states. A life history starts with "no breast cancer" before the onset of preclinical screen detectable disease. There is one preinvasive state—ductal carcinoma in situ (DCIS), and four invasive states in the model, according to the T status of TNM classification (T1a, T1b, T1c, T2+).

In the basic model used in this paper the duration in the different states follows an exponential distribution. Durations of subsequent disease states are 100% correlated so that the duration of the total preclinical screen detectable period is also approximately exponentially distributed. The transition to the clinically diagnosed states (with the same subdivision) is governed by the age-specific incidence rates and the stage distribution in a situation without screening. When cancer is detected early, women will enter the screen detected states (again with the same subdivision). Assumptions about the mean duration by age of the screen detectable preclinical states of breast cancer and the sensitivity of screening were validated using data from the Dutch screening projects in Nijmegen and Utrecht.[22] The mean duration of the preclinical screen detectable period increases with age from 1.8 years at age 35 to 6.2 years at age 70. At age 60 the mean duration times for the different preclinical screen detectable disease states are 5.2 years for DCIS, 0.14 years for T1a, 0.72 years for T1b, 1.49 years for T1c, and 1.12 years for T2+. The sensitivity of mammographic screening assumed for the different disease states is 40% for DCIS, 65% for T1a, 80% for T1b, 90% for T1c, and 95% for T2+.

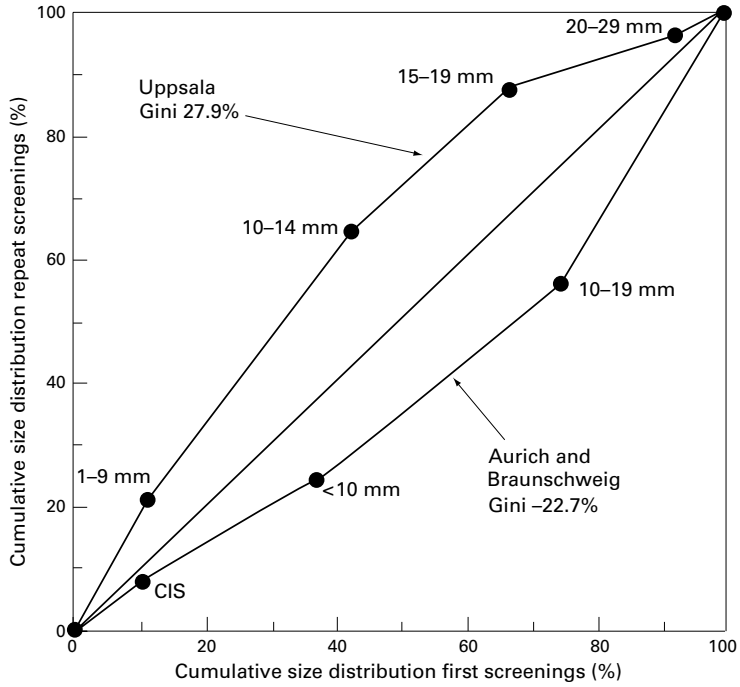

Figure 1 Illustration of the Gini coefficient for two screening projects: cumulative distribution of tumour sizes at first screenings versus subsequent screenings.

Three variants of this basic model are considered in this paper.

Variant I assumes that there is no correlation between the duration of consecutive disease states of a cancer. To maintain the same variance of the total preclinical period as in the basic model, model variant I assumes also that the duration in one disease state has a greater variance than in the basic model: a Weibull distribution with shape 0.6. This means that a relatively large proportion of cancers have a very short dwelling time in a particular disease state. For example, in an exponential distribution 5% of the cancers have a dwelling time of less than 5% of the average dwelling time, whereas for a Weibull distribution with shape 0.6 that fraction of fast growing cancers is not 5% but 15%.

Variant II assumes that tumours that will be diagnosed in a situation without screening in a later stage, have a faster growth rate than cancers which are clinically diagnosed in an earlier stage. In model variant II the duration in subsequent disease states is chosen such that cancers diagnosed as T2+ grow twice as fast as those diagnosed as T1c, while cancers diagnosed as T1c in their turn grow twice as fast as those diagnosed as T1a or T1b. The prevalence of each disease state is chosen to equal that in the basic model. This means that of those cancers that are clinically diagnosed as at most T1b, the average duration in T1a at age 60 is 0.34 years and in T1b 1.68 years; of those that are clinically diagnosed as T1c, the duration in T1a is 0.17 years, in T1b 0.84 years, and in T1c 2.18 years; and of those that are clinically diagnosed as T2+, duration in T1a is 0.09 years, in T1b 0.42 years, in T1c 1.09 years, while the duration in T2+ of course does not change (1.12 years) because they are all clinically diagnosed as T2+.

Variant III combines the adjustments to the basic model of both variants I and II by assuming no correlation between consecutive disease states. This is compensated for by a greater variance for each disease state than in the basic model: a Weibull distribution with shape 0.7 (slightly larger than in variant I) and the same average durations of disease states as in variant II.

The differences between these model variants do not affect the stage distribution in a situation without screening.

Results

Table 1 provides an overview of available data on stage distribution by tumour size at first and repeat screenings. The first column of table 1 indicates whether the data concern initial versus subsequent screenings or first versus repeat screening rounds, where the repeat screening rounds can contain some first screenings. Figure 2 summarises these data by the Gini coefficients. The Gini coefficients indicate clearly that for most screening projects, stage distribution at repeat screenings is not much better than at first screenings, and often even worse.

The two county study shows a clearly more favourable tumour size distribution at repeat screenings than at first screenings.[2 3 9 10] The trials in Malmo[1] and Stockholm[5] report similar, differences. The two reports from the Edinburgh trial do not show unequivocally whether stage distribution is more favourable at first than at repeat screenings.[7 8] Also, screening projects in Italy,[16] Germany,[11] Finland,[15] California,[14] and in Spain[17] do not show a more favourable stage distribution at repeat screening than at first screenings.

Only from the HIP trial[25] and from projects in Uppsala,[12] the UK,[18] and Ontario[19] are there reports of a substantially better stage distribution at repeat screenings in comparison with first screenings. In Uppsala there was an important increase in screening performance from the second screening round onward, which is illustrated by the fact that the detection rates in the first and second round of screening are equal. This explains why relatively few small tumours were found at first screenings.

There is no apparent correlation between screening interval and difference between stage distribution at first and repeat screenings, whereas one would expect that a longer screening interval would lead to a worse stage distribution at repeat screenings.

In summary, of the 16 screening projects from which data have been studied here, 10 show a better stage distribution by tumour size at repeat screenings, five show a worse distribution, and from one trial, one report shows a better and another shows a worse distribution.

Table 2 gives the expectations from the Miscan model for the Dutch national breast cancer screening programme. The basic model clearly expects a much more favourable stage distribution at repeat screening than at first screenings than is observed. The model predicts a Gini coefficient of 13.0%, whereas 1.2% is observed with a 95% confidence interval of –4.6% to 7.0%. Model expectations for other screening projects would be very similar to the one for the Dutch programme. Differences in model expectations would be due to differences in screening interval and ages of screened women.

Model variant I (without correlation between durations of subsequent disease states) shows a substantially less favourable expected stage distribution at repeat screening than the basic model while the distribution at first screening is very similar. The Gini coefficient is 8.0%.

Model variant II (where cancers that are diagnosed in later stages have a higher growth rate) shows similar stage distributions as variant I. The Gini coefficient is 9.0%.

The combined effects of adjustments from model variants I and II as described by model variant III further worsen the stage distribution at repeat screening, leading to a Gini coefficient of 4.0%.

The Gini coefficient from variant III and the basic variant of the model for the Dutch programme are signified in fig 2 for easy comparison of observed and model outcomes.

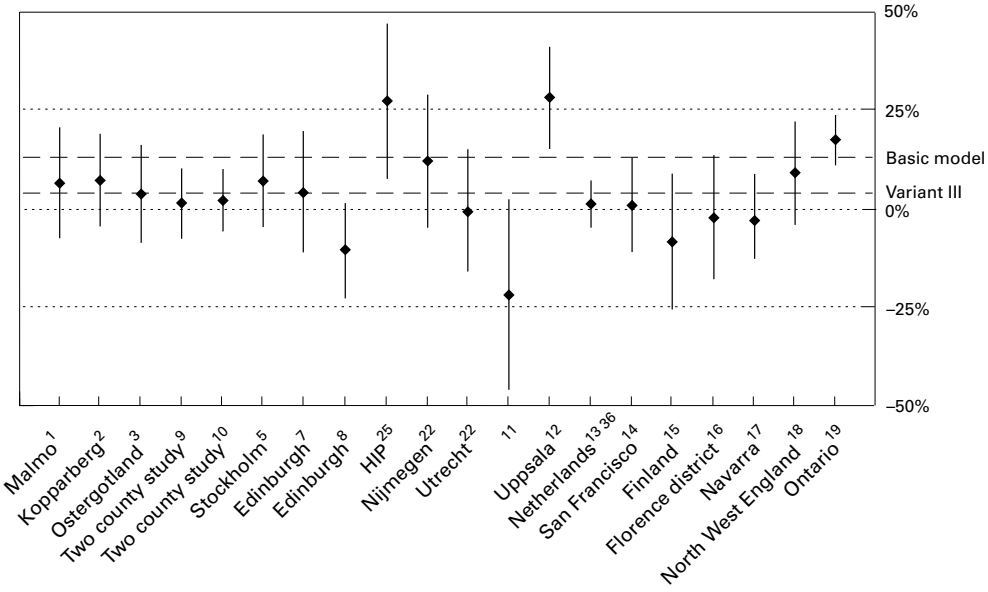

Figure 2 Gini coefficients for all screening projects, bars indicating the 95% confidence interval, horizontal dashed lines indicating model-predicted Gini coefficients.

Table 2 Expected values for the stage distribution at first and repeat screening and for the Gini coefficients, for basic model and the three model variants

	DCIS (%)	*T1a (%)*	*T1b (%)*	*T1c (%)*	*T2+ (%)*	*Gini coefficients (%)*
Stage distribution in situation without screening	4.1	1.5	6.9	33.2	54.4	
Basic model						
First screening	16.4	3.6	21.0	40.5	18.5	13.0
Second screening	14.4	6.0	33.1	37.6	8.9	
Variant I						
First screening	15.6	3.7	20.8	40.3	19.5	8.0
Second screening	13.7	6.1	28.5	37.6	14.1	
Variant II						
First screening	16.5	3.3	18.9	41.5	19.8	9.0
Second screening	13.6	5.5	28.5	39.7	12.7	
Variant III						
First screening	16.4	3.7	18.7	40.8	20.4	4.0
Second screening	12.8	5.3	25.9	39.1	17.0	

Discussion

The observed overall similarity of stage distribution at first screenings and repeat screenings is different than expected. We have tried to quantify this expectation by using the model that was validated against all data from the Nijmegen and Utrecht screening projects. Most observed Gini coefficients fall far below the expected 13%. We can conceive of four possible explanations for this discrepancy:

LITTLE CORRELATION BETWEEN DURATION OF CONSECUTIVE DISEASE STATES

The basic model assumes a large variance in rate of development of screen detectable preclinical cancer: the time spent between becoming screen detectable and clinical diagnosis follows an exponential distribution. However, given a certain duration in preclinical cancer for an individual, the proportion of time spent in each of the different preclinical disease states is fixed. This is clearly a simplification. When still assuming exponential distribution of the total screen detectable period, but no correlation between subsequent preclinical disease states, the time spent in one preclinical disease state will have a more than exponential variance. Such higher variance in duration in earlier disease states will lead to a larger proportion of cancers having grown into a larger state during the interval between screenings, thus to a less favourable stage distribution at repeat screenings.

Therefore, instead of the 100% correlation between duration in preclinical disease states as in the basic model, model variant I assumes the other extreme: no correlation. This model leads to an expected Gini coefficient of 8%, which is still higher than most observed Gini coefficients.

The correlation between duration periods, as discussed in the previous paragraphs, cannot actually be observed because it requires more than two observations of the disease state during the preclinical development of the cancer. Further research therefore is necessarily limited to theoretical approaches. Perhaps this theoretical approach in the future will lead to hypotheses that can be tested.

LATENT CANCERS

Another explanation for the fact that observed Gini coefficients fall far below the expected 13% is the presence of latent cancers, which are much more prevalent at first screenings than at repeat screenings. Latent cancers may be cancers that can be detected by screening, but which would never be clinically diagnosed in a situation without screening, or they are cancers with a very long preclinical screen detectable period. The first of these two groups would result in a non-transient increase in incidence due to screening. Such an increase is not observed (therefore is at most small) and assuming an exponential distribution of the preclinical screen detectable period as in the basic model accurately predicts observed temporary changes in incidence.[26]

The effect of latent cancers which is already in the basic model can be further increased by assuming a correlation between duration of preclinical disease and stage at clinical diagnosis.

If fast growing cancers are, on average, clinically diagnosed in a later stage than slow growing tumours, then the stage distribution at repeat screening is expected to be worse than in the basic model.

We do not know of any attempt to measure a correlation between growth rate and stage at diagnosis outside screening. However we think that the assumed ratio of growth rates of 1:1:2:4 for cancers that will be diagnosed as T1a, T1b, T1c, or T2+ as in model variant II represents a correlation at the high end of the plausible range. The expected Gini coefficient of 9% is again higher than most observed Gini coefficients. Assuming a higher correlation would imply extremely short dwelling times in disease stages preceding the stage in which the cancer is diagnosed without screening, thus a low probability of being detected at screening in an earlier stage than if there were no screening. This does not agree with the fact that the stage distribution of screen detected cancers is generally much better than that of clinically diagnosed cancers.

It is possible to investigate this correlation in studies such as those by Peer *et al*[27] or Brekelmans *et al*.[28] A correlation between growth rate and tumour size at diagnosis can be determined from the size at diagnosis outside screening and the apparent size of the tumour as it appears on one previous mammogram. Obviously, clinically diagnosed cancers that can be seen on a previous mammogram will form a selected sample, but this does not necessarily invalidate the test for correlation between growth rate and stage at diagnosis.

Even the combination of the two previously described assumed extreme model variants into model variant III, with a Gini coefficient of 4%, still leads to an expected stage distribution at repeat screenings that is more favourable than at first screenings.

FALSE REASSURANCE

A negative screening result may induce false reassurance in a woman, who may postpone a visit to the doctor for a lump she may feel in her breast, and in the doctor, who may postpone further diagnostic tests. Such a delay by the patient or doctor may lead to a later diagnosis

of interval cancers and to detection of cancers at repeat screenings that otherwise would have been diagnosed earlier as an interval cancer. For this article, only the last possibility is relevant because it influences stage distribution at repeat screenings.

The stage distribution of interval cancers is less favourable than that of screen detected cancers. A delay in the detection of such cancers up to the moment of a next screening will thus worsen the stage distribution in subsequent screening rounds. The current Miscan model cannot model false reassurance.

Delay in diagnosis of breast cancer has been studied earlier,[29–31] particularly its unfavourable effect on prognosis.[32–34] The influence on this delay of a negative finding at a preceding screening round has been reported earlier,[35] but still urgently needs to be investigated empirically.

DECREASING SENSITIVITY

The Dutch population screening programme uses double view mammography at first screenings and single view at repeat screenings. Also, it is alleged that there is a pressure on the programme to increase the number of mammographies carried out per screening unit, which might have led to a decreased quality of work since the start of the programme. These two factors may result in a lower sensitivity in repeat screenings, particularly for smaller cancers, and therefore to a less favourable stage distribution. However, it is not plausible that these effects have occurred in so many screening projects all over the world.

It is possible to investigate whether first and repeat mammograms have the same technical quality, are interpreted according to equal standards, and if two view mammography in repeat screenings will give a better sensitivity, particularly for smaller cancers. The latter has already been shown for first screenings.[36] It is also possible to investigate a possible difference in sensitivity after first and repeat screenings by comparing the incidence of interval cancers after both screenings.

Conclusion

The stage distribution at repeat examinations in most breast cancer screening projects is less favourable than expected. Four possible explanations are presented in this paper. Two of these ("little correlation between duration of consecutive disease states" and "latent cancers") have been shown to be, at best, only part of the explanation for the discrepancy between observed and expected distributions, even when both are working together; only the "latent cancers" can be empirically investigated. A decrease in sensitivity during a screening programme seems an unlikely explanation for the discrepancy found in all the screening programmes considered in this article. The reader is urged to present other explanations, but there does seem to be a serious possibility that false reassurance has a substantial effect on the observed stage distributions at repeat examinations, which are less favourable than expected.

1 Andersson I. Breast cancer screening in Malmo. *Recent Results Cancer Res* 1984;**90**:114–16.
2 Tabar L, Akerlund E, Gad A. Five-year experience with single-view mammography randomized controlled screening in Sweden. *Recent Results Cancer Res* 1984;**90**:105–13.
3 Fagerberg G, Baldetorp L, Grontoft O, *et al.* Effects of repeated mammographic screening on breast cancer stage distribution. Results from a randomised study of 92 934 women in a Swedish county. *Acta Radiologica: Oncol*ogy 1985;**24**:465–73.
4 Frisell J, Glas U, Hellstrom L, *et al.* Randomized mammographic screening for breast cancer in Stockholm. Design, first round results and comparisons. *Breast Cancer Res Treat* 1986;8:45–54.
5 Frisell J, Eklund G, Hellstrom L, *et al.* The Stockholm breast cancer screening trial—5-year results and stage at discovery. *Breast Cancer Res Treat* 1989;**13**:79–87.
6 Tabar L, Fagerberg G, Duffy SW, *et al.* The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health* 1989;**43**:107–14.
7 Roberts MM, Alexander FE, Anderson TJ, *et al.* Edinburgh trial of screening for breast cancer: mortality at seven years. *Lancet* 1990;**335**:241–6.
8 Anderson TJ, Lamb J, Donnan P, *et al.* Comparative pathology of breast cancer in a randomised trial of screening. *Br J Cancer* 1991;**64**:108–13.
9 Duffy SW, Tabar L, Fagerberg G, *et al.* Breast screening, prognostic factors and survival—results from the Swedish two county study. *Br J Cancer* 1991;**64**:1133–8.
10 Tabar L, Fagerberg G, Duffy SW, *et al.* Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* 1992;**30**:187–210.
11 Frischbier HJ, Robra BP, Hoeffken W, *et al. Mammographie in der Krebsfrueherkennung*. Stuttgart: Ferdinand Enke Verlag, 1994:314.
12 Thurfjell EL, Lindgren JA. Population-based mammography screening in Swedish clinical practice: prevalence and incidence screening in Uppsala County. *Radiology* 1994; **193**:351–7.
13 de Koning HJ, Fracheboud J, Boer R, *et al.* Nation-wide breast cancer screening in the Netherlands: support for breast-cancer mortality reduction. National Evaluation Team for Breast Cancer Screening (NETB). *Int J Cancer* 1995;**60**:777–80.
14 Frankel SD, Sickles EA, Curpen BN, *et al.* Initial versus subsequent screening mammography: comparison of findings and their prognostic significance. *Am J Roentgenol* 1995;**164**:1107–9.
15 Hakama M, Holli K, Isola J, *et al*. Aggressiveness of screen-detected breast cancers. *Lancet* 1995;**345**:221–4.
16 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;**3**:348–53.
17 van den Akker-van Marle ME, Reep-van den Bergh CM, Boer R, *et al.* Breast cancer screening in Navarra: interpretation of a high detection rate at the first screening round and a low rate at the second round. *Int J Cancer* 1997;**73**:464–9.
18 Boer R, de Koning H, Threlfall A, *et al*. Cost effectiveness of shortening screening interval or extending age range of NHS breast screening programme: computer simulation study. *BMJ* 1998;**31**7:376–9.
19 Libstug AR, Moravan V, Aitken SE. Results from the Ontario breast screening program, 1990–1995. *J Med Screen* 1998;**5**:73–80.
20 Ekelund R, Tollison R. *Economics*. Boston: Little, Brown, 1986.
21 Efron B, Tibshirani R. *An introduction to the bootstrap*. New York: Chapman & Hall, 1993.
22 van Oortmarssen GJ, Habbema JD, van der Maas PJ, *et al*. A model for breast cancer screening. *Cancer* 1990;**66**:1601–12.
23 de Koning HJ, van Ineveld BM, van Oortmarssen GJ, *et al*. Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer* 1991;**49**:531–7.
24 Habbema JD, van Oortmarssen GJ, Lubbe JT, *et al.* The MISCAN simulation program for the evaluation of screening for disease. *Comput Methods Programs Biomed* 1985;**20**: 79–93.
25 van Oortmarssen GJ, Habbema JD, Lubbe JT, *et al.* A model-based analysis of the HIP project for breast cancer screening. *Int J Cancer* 1990;**46**:207–13.
26 Boer R, Warmerdam P, de Koning H, *et al.* Extra incidence caused by mammographic screening. *Lancet* 1994;**343**:979.
27 Peer PG, van Dijck JA, Hendriks JH, *et al.* Age-dependent growth rate of primary breast cancer. *Cancer* 1993;**71**: 3547–51.
28 Brekelmans CT, van Gorp JM, Peeters PH, *et al.* Histopathology and growth rate of interval breast carcinoma. Characterization of different subgroups. *Cancer* 1996;**78**:1220–8.
29 Caplan LS, Helzlsouer KJ. Delay in breast cancer: a review of the literature. *Public Health Rev* 1992;**20**:187–214.
30 Bedell MB, Wood ME, Lezotte DC, *et al.* Delay in diagnosis and treatment of breast cancer: implications for education. *J Cancer Educ* 1995;**10**:223–8.
31 Caplan LS, Helzlsouer KJ, Shapiro S, *et al.* Reasons for delay in breast cancer diagnosis. *Prev Med* 1996;**25**:218–24.
32 Elwood JM, Moorehead WP. Delay in diagnosis and long-term survival in breast cancer. *BMJ* 1980;**280**:1291–4.

33 Neave LM, Mason BH, Kay RG. Does delay in diagnosis of breast cancer affect survival? *Breast Cancer Res Treat* 1990; **15**:103–8.
34 Porta M, Gallen M, Malats N, *et al.* Influence of "diagnostic delay" upon cancer survival: an analysis of five tumour sites. *J Epidemiol Community Health* 1991;**45**:225–30.
35 Burns PE. False-negative mammograms delay diagnosis of breast cancer. *N Engl J Med* 1978;**299**:201–2.
36 Blanks RG, Moss SM, Wallis MG. A comparison of two view and one view mammography in the detection of small invasive cancers: results from the National Health Service breast screening programme. *J Med Screen* 1996;**3**:200–3.