Volume 43, Number 3, 1990 ISSN 0895-4356

Journal of Clinical Epidemiology

Formerly Journal of Chronic Diseases

CONTENTS

J Clin Epidemiol is indexed/abstracted in Curr. Cont. *ASCA*, Biosis Data., BRS Data., CAB Inter., Cam. Sci. Abstr., Curr. Cont./Cli. Med., Curr. Cont./Life Sci., CABS, Excerp. Med., High. Educ. Abstr., Curr. Cont. ISI/BIOMED Database, Soc. Ibe. de Inf. Cie., Curr. Cont. Sci. Cit. Ind., Curr. Cont. SCISEARCH Data., Ind. Med.

PERGAMON PRESS

Oxford · New York · Beijing · Frankfurt · São Paulo
Sydney · Tokyo · Toronto

ISSN 0895-4356

JOCDAE 43(3) 211–312 (1990)

J Clin Epidemiol Vol. 43, No. 3, pp. 215–225, 1990
Printed in Great Britain.

0895-4356/90 $3.00 + 0.00

Variance and Dissent

THE EPIDEMIOLOGY OF MASS BREAST CANCER SCREENING—A PLEA FOR A VALID MEASURE OF BENEFIT

JOHANNES G. SCHMIDT

Centre for Clinical Epidemiology, University of Newcastle, NSW 2308, Australia
and Institut Universitaire de Médecine Sociale et Préventive, CH-1005
Lausanne, Switzerland

(*Received in revised form* 1 *August* 1989)

Abstract—The present paper analyses the epidemiologic effects of mass breast cancer screening. Mass mammography may possibly achieve a breast cancer mortality reduction in relative risk terms. However, this does not necessarily represent a net benefit. It is argued that the benefits and adverse effects of a screening programme must be measured in terms of absolute risks. According to this measure, the mortality reduction achieved by a mass breast screening programme is only one death per approx. 15,000 women-years. Many thousands of mammograms are needed to prevent one cancer death, and for each woman who can derive a direct benefit in terms of a prevented breast cancer death, hundreds of women have to suffer the anxiety of a positive screening mammography. Moreover, it is possible that adverse effects of breast cancer screening may contribute to mortality from other causes. Even with the assumption that screening can save lives, the net health effect of mass breast cancer screening is questionable and appears to be rather detrimental. It may be an error to recommend mass breast screening.

Mass screening Breast cancer Mammography Critical appraisal

INTRODUCTION

The breast cancer screening problem is often reduced to one question: does screening and early intervention (mastectomy/lumpectomy) truly reduce the mortality from breast cancer? From a more comprehensive epidemiological point of view, however, the problem must include the question: do benefits outweigh the adverse effects and in what terms should both be assessed?

All correspondence should be addressed to: J Schmidt, M.D., Gesundheitsplan HMO (Swiss HMO Health Plan), Group Practice Schmiedhof Zweierstrasse 136/138, CH-8003 Zurich, Switzerland.

NATURAL HISTORY OF BREAST CANCER

Historically, breast cancer screening is based on the premise that this cancer spreads in an orderly manner from a local focus to regional lymph nodes and only then disseminates to distant tissues. It is assumed that there is a "critical point" in the natural history of this disease before which therapy is more effective and easier to apply than afterwards. It is also assumed that for a period of time after early diagnosis becomes possible and before the critical point, treatment may arrest disease progression ("window period"). However, we do not know where the critical point and how long the

window period for early intervention exactly is. Available evidence from screening trials suggests that detection by screening provides the potential of a cure in rather advanced cancers with node involvement, more than in early cancers [1]. Some have hypothesized that many breast cancers act as a systemic cancer from inception [2, 3]. Microscopic cancer foci have been found in the bone marrow of a considerable proportion (15%) of patients with early cancer [4], and presumably, early detection and local treatment in such patients would not be curative. Further, the clinical nature of many "minimal" cancers may be benign [5]. ("Minimal cancer" is synonymous with cancer *in situ* or very small lesions ($\leqslant 0.5$ cm), "early cancer" with stage I cancer ($\leqslant 2$ cm, no node involvement) [6].)

Early detection may be curative in a subset of breast cancers. Early detection, however, is not synonymous with cure and does not necessarily convey a benefit to each patient.

SCREENING TRIALS

Only randomized controlled trials are qualified to yield an unbiased measure of the effects of breast cancer screening because of the "lead-time bias", the "length bias", the "healthy-screenee effect", and the "overdiagnosis bias" that are involved in other study types [7, 8]. Case–control studies are useful to some extent, but they overcome neither the healthy-screenee effect nor the possible bias that deaths from prognostically unfavourable, fast growing, cancers will necessarily constitute the majority of the "cases"; these are less likely to be detected by screening (length bias), and this means that women with fast growing cancers also have a lesser chance of participating in a screening programme. The repeatedly observed two-fold overestimation of the relative reduction in breast cancer mortality in case–control studies [9–11] is consisent with such inherent biases. Although case–control studies may be able to demonstrate a true breast cancer mortality reduction, they are not qualified to measure the true extent of this effect.

Mortality results from three randomized controlled trials, from the New York HIP study and from two Swedish studies, the "two-county" and the Malmö trial, are available. Further, mortality results from the U.K. trial of early detection of breast cancer have recently been published [12]. This study, although with a prospective design, was not randomized, thus subject to uncontrolled biases. More randomized trials involving mammography are under way in Canada and in Edinburgh, as well as in other Swedish towns. A randomized study of breast self-examination is in progress in Leningrad and Moscow [13].

The two-county trial, however, is somewhat difficult to evaluate. The results in subsequent publications are given for continually changing time-periods and sub-populations (initially all women above 40 were enrolled), and blinding and confirmation procedures for the determination of end points appear to be poorly defined. Randomization took place at community level and not at individual level, and thus, some bias due to possible baseline differences, socioeconomic for example, cannot be excluded. The Malmö trial appears free of such problems and could therefore be regarded more reliable than the two-county study. The early HIP trial does not reflect the currently applied technology and, therefore, does not form an integral part of the evaluation of screening in this paper.

For this reason, the evaluation of the effects of breast cancer screening will mainly be based on the two Swedish trials (i.e. Kopparberg/Östergötland and Malmö).

SCREENING: PERFORMANCE, CANCER YIELD, AND INVESTIGATION LOAD

The radiation hazard of modern mammography screening is very small, the proportion of breast cancers induced by radiation being 1% at the most (based, however, on indirect estimations from nuclear radiation and not on direct observation) [14–16]. Modern mammography is considerably more sensitive and involves lower levels of radiation than the technology used in the earlier HIP study.

The Swedish trials demonstrate the potential of modern mammography screening. In the study conducted in the two-counties Kopparberg and Östergötland, 65% of all cancers in the screening group were early cancers compared to approx. 40% in the control group ($p < 0.001$). Single-view mammography was the only screening method used and was scheduled for every 24 months in the younger age groups (< 50 years) and every 33 months in older ages [17]. In Kopparberg, the participation rate was excellent—about 90% (women of age 40–69) [18]. In the Malmö study, using two-view mammography at 18–24 months screening intervals, an

Table 1. Sensitivity and negative predictive value of a breast cancer screening programme. Two-county study, 7 years follow-up, all rounds combined

Sensitivity	
On screening*	75%
Intention-to-screen†	67%
Negative predictive value‡	99.6%
Pre-test likelihood of no disease§	98.5%

*Missed cancers = interval cancers alone.
†Missed cancers = interval cancers + cancers in refusers.
‡Chance of not having (interval) cancer with a negative result.
§Chance of not having cancer without screening test.

equally successful detection of early stage cancer in the entire target population was achieved (68 vs 48%, $p < 0.001$) notwithstanding a compliance of "only" about 70% [19]. This shows that, at least in a society such as the Swedish one, mass mammography screening is feasible in practical terms.

In the two-county study, 1191 invasive breast cancers emerged in the screened group ($n = 78{,}085$) during 7 years follow-up. 797 were detected by mammography, 261 cancers developed during a screening interval ("interval cancers") and 133 cancers appeared in women who refused to participate [20]. The sensitivity of the screening programme and the negative predictive value of an individual screening result can be derived from these figures (Table 1).

As the compliance in the two-county trial was high the sensitivity of the programme in the overall target population was also high (67%). An almost identical "intention-to-screen" sensitivity was achieved in the Malmö trial (64%; "on-screening": 79%, figures derived from [19]). It can be seen, however, that even with a technically excellent screening programme one third of breast cancers could not be detected earlier with screening than without such a programme, although more frequent examinations could increase the sensitivity to some extent [21]. The 75% sensitivity of the two-county trial, in compliers, is comparable to that of approx. 75% in the BCDDP project [22] and to the 69% in the ongoing Canadian trial [23], both using annual mammography.

The negative predictive value of a mammography screening result is 99.6% for all rounds combined and 99.8% for the initial screening round (latter figure derived from the Kapparberg study [18]). A woman with a negative result, thus, can be assured with a certainty of close to 100% of neither already having nor developing cancer within the next screening interval. However, irrespective of a mammography result, the overall risk that a participating woman would have breast cancer during the reported 7 years was only 1.5 and 0.6% for the first screening round and interval (latter figure derived from [18]). It can be argued that, in individual terms, it makes no difference whether a woman can be 99.8% certain, rather than 99.4% certain, of being free of cancer during a screening interval. Therefore, screening appears to offer more false reassurance than real certainty.

During 7 years from the start of the trial (average 6.25 years from randomization) the activities shown in Table 2 took place in Kopparberg [18].

From these figures the specificity of the initial screening mammography can be calculated as 92% (all rounds together; the corresponding figure for the first round alone, with a higher proportion of more advanced unequivocal cancers, is 95%). A slightly higher specificity of 97% was achieved in Malmö, reflecting the use of two views (figure derived from [24]). In comparison, a 94% specificity has been reported for the first round in the Canadian trial [23].

Of greater direct practical interest than the specificity is the positive predictive value which can be estimated for the various diagnostic steps. For example, 9.7 positive single-view mammograms per confirmed case correspond with a positive predictive value of 10.3% (i.e. 1/9.7), or 3.3 positive complete mammographies per case with a value of 30%. The results of the Malmö trial are similar with a positive predictive value for the screen of 14% and of 37% for complete mammography [24]; the positive predictive value in the Canadian trial was reported 8.6% [23]. In other words, only about 10% of initially positive screenees were eventually diagnosed as having cancer; 90%, who did not have cancer, had to undergo an additional complete mammography and possibly other types of investigation. The Swedish trials achieved a very

Table 2. Breast cancer screening: screening activities per 10,000 invited women in 7 years; Kopparberg Study*

	Total number	Per case detected
Single-view mammography	20,100	248
Complete mammography	790	9.7
Needle biopsy/clinical examination	270	3.3
Surgical biopsy	140	1.7
Early detected cancer	80	—

*Reported for ages 40–69, $n = 33{,}641$.

Table 3. Breast cancer incidence* and stage distribution per 1000 in study and control group in the two Swedish studies

	Total cancers†	Stages II–IV	Node positive/ disseminated
Kopparberg/Östergötland (age 40–74 at entry)			
Study ($n = 78,085$)	16.7 (100%)	5.9 (35%)	4.1 (25%)
Control ($n = 56,782$)	13.5 (100%)	8.0 (59%)	5.0 (37%)
Relative risk	1.24	0.74	0.81
Difference	−3.2	2.1	0.9
Significance	$p < 0.001$	$p < 0.001$	$p = 0.011$
Malmö (age 45–69 at entry)			
Study ($n = 21,088$)	27.9 (100%)	9.0 (33%)	
Control ($n = 21,195$)	21.1 (100%)	10.9 (52%)	
Relative risk	1.32	0.83	
Difference	−6.8	1.9	
Significance	$p < 0.001$	$p = 0.5$	

*Average follow-up of 7.5 years in Kopparberg, 6.5 years in Östergötland, and 8.8 years in Malmö.
†Includes *in situ* cancer.

good screening performance. In the ongoing Edinburgh Breast Screening Project, 3 biopsies were necessary per confirmed cancer case [25], which represents a positive predictive value for the pre-biopsy examinations of 33%. This can be compared to 59% in Kopparberg county (i.e. 1/1.7). Also, preliminary data from the Canadian trial suggest that more biopsies may be needed per definite cancer diagnosis; among non-palpable tumours only 1 in 10 biopsies resulted in a confirmed cancer [8].

Obviously, the "burden of investigation" may become considerably greater than suggested by the above results from Sweden, if very sensitive screening methods, such as annual two-view mammography, are applied, or if biopsy is used more generously.

The cancer yield of the Swedish screening programmes [19, 21] is presented in Table 3.

The figures in Table 3 confirm the ability of the Swedish mammography programmes to detect more cancers earlier. It can be argued that the reduction of advanced cancer (i.e. stages II–IV), consistent in both studies, represents a gain in terms of a decrease in radical surgery (simple lumpectomy instead of mastectomy for early cancer). At the same time, however, significantly more clinically detected cancers occurred in the screening groups. In fact, the screening efforts in Kopparberg led to a significant 1.9-fold increase of inpatient breast surgery throughout the entire trial period; approx. 260 operations on an average were performed annually during the entire trial compared to 140 annual operations in the pre-trial period. In a non-randomized control county, simultaneously, the corresponding increase was only 1.4-fold [26]).

To some extent, this cancer case excess in the screening group may be due to a desirable lead-time effect, and thus, the difference in overt cancer between screening and control group may perhaps diminish in the end. The 10-year cancer incidence in the HIP trial may suggest an eventual equalization of the breast cancer incidence between the study and control group [27]. The sensitivity of the HIP screening programme, however, was comparatively low, and its results do not allow conclusions about the effects of modern mammography. In the Malmö trial, using more sensitive mammography, the cancer case excess after 9 years was over 30% (Table 3). It appears likely that modern screening detects some breast cancers which otherwise would not have manifested themselves clinically. The available data suggest that many cancers may be clinically benign. A Danish autopsy study [5] of 83 women of a median age of 67 years at death (range: 22–89 years) showed that 21 (25%) had some form of breast carcinoma of which only 6 (29% of the cancers) had been clinically apparent before death. Therefore, as many as 70% (95% confidence interval: 52–90%) of breast neoplasms never developed into a clinically relevant disease. These breast neoplasms, presumably, are potentially detectable by mammography. Moreover, from the BCDDP project it is known that false positive histology results may occur quite often (in 17% of minimal cancers and in 6% of screening-detected cancers altogether [28]). Presumably, overdiagnosis is all the more likely the more sensitive a screening is in detecting minimal cancer. An increasing proportion of early, minimal cancer with a questionable "clinical malignancy" and a sometimes

equivocal histology would be detected in a continuing screening programme. Overdiagnosis, presumably, would thus increase in the end. Due to the fact that no active trial phase has lasted longer than 9 years, there is a lack of conclusive data to assess lead-time effect and overdiagnosis in the long-term perspective. As the radiation hazard of modern mammography is very small (causing $\leqslant 1\%$ of breast cancers [14–16]), and because of a latency period of about 10 years, possible radiogenic breast cancer effects did not contribute to the observed cancer case excess.

Practically, the above given "cancer yield" implies that approx. 2 out of 1000 women in 7–9 years could enjoy the benefit from lumpectomy rather than needing mastectomy, but about 5 women, simultaneously, became cancer cases and were subjected to surgery that would not have been performed without screening. The U.K. trial even showed a 51% higher rate of diagnosed breast cancer in the population invited to screening, accounting for 9 additional cancer cases per 10,000 women-years [12].

EXTENT OF BREAST CANCER MORTALITY REDUCTION

Crude breast cancer mortality figures from the two Swedish trials [19, 21, 29] are given in Table 4.

The observed breast cancer mortality reduction of 29% (i.e. relative risk = 0.71) in the two-counties of Kopparberg and Östergötland is statistically significant. However, no significant breast cancer mortality reduction could be observed in the recently published Malmö trial (consistent with the negative findings of the non-randomized prospective U.K. trial using annual mammography plus clinical examination (relative risk = 0.83, $p = 0.06$) [12]). It is difficult to interpret this discrepancy [19]. Although the protocol of the two-county trial does not exclude biases, it may be inappropriate to ignore this trial and to consider the Malmö study alone.

In this paper, a pooled analysis of both studies will be used to estimate the extent of the breast cancer mortality reduction that may be achievable by screening. The breast cancer mortality reduction shown by such an analysis is statistically significant (Table 4) and it might be concluded, therefore, that mammography screening may reduce breast cancer mortality. This would suggest that a considerable subset of breast cancers are of a "local type", in which the removal of the locally apparent cancer focus may prevent or delay death from breast cancer. This result may therefore be significant for the natural history concept of breast cancer. (The clinical significance of this result, however, is another question and will be discussed below.)

The two-county trial suggests a greater mortality reduction of 12 deaths per 100,000 women annually if women under the age of 50 at entry are excluded (50–74 age group), compared to 8 per 100,000 annually in the entire group (40–74 age group) [29]. A (non-significant) breast cancer mortality reduction for

Table 4. Breast cancer mortality in the Swedish trials. Average follow-up of 8.8 years in Malmö and of 6.5 years in the two counties

	Breast cancer deaths	Rate per 10,000	Difference* per 100,000 women-years	Relative risk
	Kopparberg/Östergötland†			
Screening group ($n = 78{,}085$)	104	13.3		
40–74 Age group			8.2 (1.5; 15)	0.71[a]
Control group ($n = 56{,}782$)	106	18.7		
50–74 age group§			12.1[b]	0.61[b]
	Malmö‡			
Screening group ($n = 21{,}088$)	63	29.9		
45–69 Age group			1.4 (−11; 23)	0.96[c]
Control group ($n = 21{,}195$)	66	31.4		
55–69 Age group			7.8	0.80[d]
	Both trials combined			
Kopparberg/Östergötland + Malmö			6.2	0.80[e]

*In brackets: 95% confidence intervals.
†Source: [21, 29].
‡Source: [19].
§6 years follow-up.
[a] $p = 0.01$ (Mantel–Haenszel odds ratio); [b] $p = 0.003$ (Mantel–Haenszel odds ratio, difference adjusted); [c] $p = 0.8$ (NS); [d] $p = 0.3$ (NS); [e] $p = 0.04$ (Mantel–Haenszel odds ratio).
Underscored values are those that are statistically significant.

women below 50 was observed in the HIP trial [30]. For women aged 40–49 at entry in the two-county study [29], as well as for women below 55 years of age in the Malmö trial [19], however, there was a higher breast cancer mortality rate in the screened group. In the Malmö trial, the statistical significance of this mortality increase is certainly poor, but almost as great as the significance of the mortality reduction in women of age $\geqslant 55$ (p-values of 0.37 and 0.31 respectively). It may also be revealing to consider that the observable increase in cardiovascular mortality in the screening group of this study was considerably more significant ($p = 0.16$) than any breast cancer mortality reduction in any sub-group. These considerations may caution against putting any emphasis on subgroup results of the Malmö trial.

Self-examination is not likely able to prevent death from breast cancer. It is questionable whether this screening method is effective at all in mere terms of early detection [13, 31]. In terms of a mortality reduction, no effect of self-examination could be shown in the large prospective (non-randomized) U.K. trial (relative risk = 1.04) [12]. Definite evidence will be generated by a randomized trial.

FALLACIES IN CONVENTIONAL MEASURES OF BENEFIT

Quite generally, the breast cancer mortality reduction in relative risk terms, as observed in the HIP and Swedish trials, is considered conclusive evidence for the net benefit of breast screening [32–36]. However, as shown in Table 4, the 20% breast cancer mortality reduction in the Swedish trials meant, in absolute risk terms, a reduction of only 6 deaths in 100,000 women-years (i.e. 1 death per approx. 15,000 women-years). This should be contrasted with the number of interventions necessary at the same time. In Kopparberg, 790 suspicious mammograms per 10,000 women resulted in 7 years screening; 270 tumours, after complete mammography, needed further investigations including 140 surgical biopsies (Table 2).

Consequently, 7000 screening mammographies were needed for the prevention of one breast cancer death. For each woman who derived a direct benefit in terms of a prevented breast cancer death, over 200 women suffered anxiety due to an initially positive screening mammogram, and approx. 30 of these women were told the diagnosis earlier than had the tumour be discovered accidentally. Expressed in terms of a "positive predictive value in predicting benefit from cure" the predictive value of a positive screening mammogram is a mere 0.4% (i.e. 1 prevented death per 230 positive mammograms). Moreover, about 10 additional cancers (cases that would not have become clinically overt without screening) were diagnosed per each breast cancer death prevented (as can be derived from the figures in Tables 3 and 4).

It can be argued that the mortality effect achieved by screening is diluted by an incomplete attendance and that the potential mortality gain for an individual willing to comply with a screening programme, therefore, may be greater than suggested by the above results. A separate (pooled) analysis of the subgroup of attenders in the two Swedish studies would show a mortality reduction of approx. 50%, accounting for about 15 prevented deaths per 100,000 women-years. This figure, questionable because of a possible healthy-screenee effect involved in such an analysis, however, does not appreciably change the picture.

Total mortality of the screened group in the two-county trial was not different from the control group [37]. It is difficult, however, to demonstrate a reduction in overall mortality due to a breast screening programme because the contribution of breast cancer mortality to overall mortality is very small. In both Swedish trials, the proportionate breast cancer mortality was only 3%; a 20–30% breast cancer mortality reduction represents a reduction of less than 1% of the total mortality. A statistically significant effect even of the most effective breast cancer screening on total mortality, thus, can never be expected.

The effect of screening on mortality due to causes other than breast cancer for the subgroup of breast cancer patients in the two-county study [37] is shown in Table 5.

The absolute case fatality from causes other than breast cancer was significantly greater in the screened group (relative risk = 1.62, $p = 0.02$). Thus, a greater number of breast cancer patients in the screened group died from causes other than breast cancer compared to the control group. The overall absolute case fatality in patients with breast cancer is similar in the study and control groups ($p > 0.4$). This can partly be explained by the fact that more diagnosed cancer cases occurred in the screened group, but the relative non-breast cancer

Table 5. Mortality among breast cancer patients in the Swedish two-county study, average follow-up of 7 years*, women aged 40–74 at entry

		Deaths among breast cancer patients		
		Non breast cancer deaths		
County		absolute per 10,000†	relative per 100 cancers	Total deaths absolute per 10,000†
Kopparberg	Study	12.0	7.0	30.2
	Control	8.0	4.9	35.6
Östergötland	Study	8.7	5.5	22.3
	Control	5.0	4.1	22.7
Combined (Mantel–Haenszel)		OR = 1.62 $p = 0.02$	OR = 1.40 $p = 0.11$ (NS)	OR = 0.92 $p > 0.4$ (NS)

*7.5 years for Kopparberg, 6.5 for Östergötland.
†Denominator = all women (not only cancer cases).

mortality among cancer patients was still 40% greater in the screened group. This latter difference, though observable in both counties, is not statistically significant; however, the likelihood of a type II error (i.e. missing a difference where in fact there is one) is well over 50%. The Malmö trial which was unable to show a breast cancer mortality reduction through mammography screening, however, did not show such an effect of screening on the case fatality in patients with breast cancer [19]. In the light of these latter results, the above findings may indicate possible classification biases in the two-county trial.

Another explanation is possible. When analysed in terms of absolute risks it is evident that a programme with such an immense "burden of investigation" may result in appreciable iatrogenic morbidity. It is even possible that subtle adverse mortality effects, however difficult to classify, could annul the small reduction in breast cancer mortality due to screening, in terms of absolute risks. It may be conjectured that adverse effects associated with a cancer diagnosis, for example the stress of the implications of the diagnosis itself or the stress of the attendant surgery, may lead to depression with an increased susceptibility to disease, or to accidents and suicide. In the screened group, women with a breast cancer diagnosis are significantly increased in numbers and exposed to a longer overt disease period. In Malmö, for example, the increase in breast cancer years through screening was 1.5-fold [19], accounting for a difference of over 500 disease years per 100,000 women-years. Thus, approx. 90 additional disease years (years with cancer that would not have been diagnosed without screening) were necessary for one prevented breast cancer death.

Thus far, there are no data to prove that breast screening prevents significantly more deaths than it eventually may cause.

COST AND BENEFIT IN COMPARISON

A meaningful cost-benefit analysis should, above all, be based on an "epidemiological cost-benefit analysis". That is to say, health benefits should clearly outweigh health costs in order to make a programme worthwhile.

Estimations of the direct financial costs of screening have been made for the two-county trial [38], but they are incomplete because they do not include the costs incurred from the cancer case excess and the consequent increase in surgery that was observed in the same study [26] (see also Table 3). Morbidity costs, i.e. loss of working hours, of mammography screening are considerable. Assuming that one mammography produces absenteeism of one half of a working day, approx. 13 years of absenteeism would result per breast cancer death prevented, given the attendance rate and screening intervals of the Swedish trials. In addition, the prevention of one breast cancer death is achieved at the expense of hundreds of initially positive mammograms and of a lengthening of the disease period in 30 patients. The time directly used for follow-up procedures and possible absenteeism attributable to stress and anxiety must therefore be considered.

The present paper, however, focuses on epidemiological benefits and costs, i.e. benefits and costs in health terms. In addition to the effects already analysed in this paper, "intangible costs" such as a constant worry of possibly having cancer [39] should be included. No respective data are available from the Swedish studies. According to the Edinburgh trial,

Table 6. Beneficial, adverse, and cost effects* of mammographic breast cancer screening. Based on the two Swedish studies

Kind of effect or cost	Number of effects/cost per 100,000 women years		
	Two county study	Malmö	Both studies combined†
Prevented deaths from breast cancer	8.2 (1.5–15)	(1.4) ($p = 0.81$)	6.2 ($p = 0.04$)
Lumpectomy instead of mastectomy (decrease of stage II+ cancer)	30 (17–43)	22 (0–43)	29 ($p < 0.001$)
Prevented overall death	?	?	?
Cancer case excess (increase in overt cancer)	46 (27–64)	77 (44–111)	52 ($p < 0.001$)
Mammograms	38,000	40,000	
Positive screening mammography	1500	1200	
Clinical examinations/cytology	690	480	
Lengthening of disease period (earlier detection by screening)	150	200	

*Figures approximate, in brackets: 95% confidence intervals.
†Differences derived using Mantel–Haenszel odds ratio.

psychiatric morbidity produced by breast cancer screening among complying women may be considered a minor problem (8% of participating women feel more anxious about breast cancer after screening [40]). However, many women (over 30% in the same study) appear not to comply with screening advice from the beginning because of the anxiety a mere screening invitation can cause [41]. Such health costs may be considered negligible unless compared to the benefits in absolute risk terms: less then 1% of women can ever expect a benefit from screening (i.e. a prevented breast cancer death).

If both, beneficial and adverse effects are measured in absolute rather than in relative risks, a comparison of the magnitude of advantages and disadvantages of screening becomes possible. These epidemiological effects, as can be observed in the Swedish studies, are summarized in Table 6.

The figures in Table 6 allow a direct comparison of the investigation load with the number of breast cancer deaths prevented. For example, 690 positive results after complete mammography compared to 6.2 prevented deaths from breast cancer, both per 100,000 women years, means that 1 of 111 women who needed needle biopsy and further surgical investigations could expect a benefit in terms of a cure. The table shows that a considerable number of women suffer adverse effects such as a lengthening of the disease period, but only a very small number of women can benefit from a cure. It should be emphasized that we do not know exactly whether mass screening prevents significantly more deaths than it ultimately may cause.

CONCLUSION

The two Swedish studies are randomized controlled trials, and are thus the only trials to date which allow an unbiased evaluation of modern breast cancer screening, including an epidemiological cost–benefit assessment. However, the trial with the least possible biases (Malmö) showed no mortality reduction and the pooled estimation of the breast cancer mortality reduction used in this analysis may thus be rather optimistic. Moreover, the Swedish trials showed an excellent performance in terms of sensitivity and specificity, in spite of a comparatively simple and not very intensive screening scheme. In general, the "burden of investigation" and associated health costs may be higher.

The clinical performance of an intervention programme, evidently, is not the most important standard; epidemiological methods in quantifying the overall balance of beneficial and adverse effects should rather be the core standard for our appraisal of medical interventions. The application of absolute risks as a measure to compare these effects is a useful method.

Although a 20% reduction of breast cancer mortality through mass mammography screening appears promising, it is a solution that will not decrease the overall burden of illness and suffering. Mass screening in the entire, unselected population will always be associated with relatively low positive predictive values, even when the tests are very specific. Moreover, the ability to cure a disease is limited, in spite of early detection. This further reduces the "positive predictive value" in predicting

eventual benefit to very low values. Only 0.4% of women with a suspicious screening mammogram, and 0.9% with a tumour suspicion after complete mammography, could profit from the prevention of a breast cancer death. Further technological improvements in screening sensitivity, thus, will not likely offer an advantage. On the contrary, they may actually increase adverse effects, such as the observed cancer case excess. The above balance, actually, is not unexpected for those who understand the fallacies of drawing inferences from mortality reductions in relative risk terms alone. In absolute risk terms, even the very optimistic 40% breast cancer death reduction in the 50–74 age group in the two-county trial represents only 1 case in 10,000 woman years. This gain is considerably less than many of the daily risks; for example, the excess mortality of a construction worker in comparison to a university professor is 135 per 10,000 annually, and the traffic mortality risk for a car driver driving an average of 14 km daily is 8 per 10,000 annually (England, 1982) [42].

The breast cancer screening trials are of theoretical interest for the elucidation of the natural history of breast cancer. They may support the validity of the local growth model, which appears to apply to a subset of breast cancers. The conventional assessment of the practical benefit of breast cancer screening (and of many other medical interventions), based on relative mortality reductions, however, is misleading and may now be considered an error. I believe that the small benefit of mass breast cancer screening does not justify the epidemiologic cost, as given in Table 6. This view is strengthened by the fact that it has not been shown that mass mammography screening is able to prevent significantly more deaths than it ultimately may cause. Breast cancer screening does likely more harm than good not only to women under age 50 years, but to women of the 50–75 age group as well! Eventually, it should be mentioned that a similar investigation into the effects of other interventions (surgery, chemotherapy) may lead to more cautious conclusions about the benefit of various breast cancer treatments (often designed as tertiary prevention) as well.

The extent of the mortality reduction in the Swedish studies is interesting when compared to the results of the New York HIP trial [1, 27, 43, 44]. The much greater sensitivity of the Swedish trials in detecting early cancer resulted in a smaller mortality reduction ("intention-to-screen" sensitivity of 67% in the Swedish trials compared to 44% in the HIP study, and a breast cancer mortality reduction of 20% in Sweden (Table 4) compared to 40% in the HIP study [43]). This is inconsistent with the concept that the effectiveness of screening is due to its ability to detect pre-clinical minimal cancer, which is supposed to be more responsive to treatment [45]. This paradox, however, may be consistent with an unexpected result from the HIP trial: beginning with the fifth year after diagnosis the screening benefit was more pronounced in axillary node positive patients than in node negative ones [1]. Thus, the critical point for early intervention may be either too early to be detectable for the incurable cancer sub-population of the "systemic type", or relatively late for the curable sub-population of the "local type" (obviously including stages with axillary node metastases). This could explain why a dramatic increase in the screening programme sensitivity did not result in a greater cure rate.

The cost–benefit ratio of a screening programme, presumably, changes with a variety of factors. The "burden of investigation" and associated health costs are reduced with an increasing positive predictive value of the screen and the diagnostic work-up steps, and the absolute benefit increases with the breast cancer mortality risk of the population screened. Screening high-risk women, therefore, may probably be beneficial. However, only a specific screening trial in a high-risk population would allow a definite evaluation of the screening benefit in this group of women. Further, the comparison of the HIP trial with the Swedish Study may suggest that the increase of the screening sensitivity beyond a certain point may adversely affect the cost–benefit ratio.

Economic analysis, assessing gains and loss in money terms, is close to the epidemiological concept of absolute attributable effects that has been developed in the present paper. The necessity of the comparison of absolute rather than relative effects has been suggested by others [46] but an appraisal of breast cancer screening in the terms developed in this paper was not published earlier. Only one recent paper, assessing the value of screening in women under age 50 years [47], uses a similar approach.

Acknowledgements—This paper was possible due to an Educational Grant from the Swiss National Fund for Research, and due to the courtesy of the University of Newcastle/Australia in admitting me as an overseas Postgraduate at the Centre for Clinical Epidemiology during 1986. I thank Mrs Jody Staehelin, MD, for corrections of the English.

REFERENCES

1. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selection, follow-up, and analysis of the Health Insurance Plan Study: A randomized trial with breast cancer screening. **Natl Cancer Inst Monogr** 1985; 67: 65–74.
2. Henderson C, Canellos GP. Cancer of the breast—the past decade (first of two parts). **N Engl J Med** 1980; 302: 17–30.
3. Henderson C, Canellos GP. Cancer of the breast—the past decade (second of two parts). **N Engl J Med** 1980; 302: 78–90.
4. Mansi JL, Berger U, Easton D *et al.* Micrometastases in bone marrow in patients with primary breast cancer: evaluation as an early predictor of bone metastases. **Br Med J** 1987; 295: 1093–1096.
5. Nielsen M, Jensen J, Andersen J. Precancerous and cancerous breast lesions during lifetime and at autopsy. **Cancer** 1984; 54: 612–615.
6. Council of Scientific Affairs. Early detection of breast cancer. **JAMA** 1985; 252: 3008–3011.
7. Sackett DL, Haynes RB, Tugwell P. **Clinical Epidemiology—A Basic Science for Clinical Medicine.** Boston, Mass: Little, Brown and Co.; 1985: 146.
8. Miller AB. Screening for cancer: issues and future directions. **J Chron Dis** 1986; 39: 1067–1077.
9. Coilette HJA, Day NE, Rombach JJ, DeWaard F. Evaluation of screening for breast cancer in a non-randomized study (the DOM Project) by means of a case–control study. **Lancet** 1984; I: 1224–1226.
10. Verbeek ALM, Hendriks JHCL, Holland R, Mravunac M, Sturmans F, Day NE. Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegan Project 1975–81. **Lancet** 1984; I: 1222–1224.
11. Palli D, Del Turco MR, Buiatti E, Carli S, Ciatto S, Toscani L *et al.* A case–control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). **Int J Cancer** 1986; 38: 501–504.
12. U.K. Trial of Early Detection of Breast Cancer Group. First results on mortality reduction in the U.K. trial of early detection of breast cancer. **Lancet** 1988; II: 411–416.
13. Day NE, Chamberlain J. Screening for breast cancer: workshop report. **Eur J Cancer Clin Oncol** 1988; 24: 55–59.
14. Fitzgerald M. Radiation hazards of breast screening. **Br J Radiol** 1983; 56: 283–284.
15. Zuur C, Broerse JJ. Risk- and cost-benefit analysis of breast cancer screening programs derived from absorbed dose measurements in The Netherlands. **Diagn Imag Clin Med** 1985; 54: 211–222.
16. Gohagan JK, Darby WP, Spitznagel EL, Monsees BS, Tome AE. Radiogenic breast cancer effects of mammographic screening. **J Natl Cancer Inst** 1986; 77: 71–76.
17. Tabár L, Gad A. Screening for breast cancer: the Swedish Trial. **Radiology** 1981; 138: 219–222.
18. Tabár L, Gad A, Holmberg L, Ljungquist U. Significant reduction in advanced breast cancer—results of the first seven years of mammography screening in Kopparberg, Sweden. **Diagn Imag Clin Med** 1985; 54: 158–164.
19. Andersson I, Aspegren K, Janzon L *et al.* Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. **Br Med J** 1988; 297: 943–948.
20. Fagerberg CJG, Tabár L. The results of periodic one-view mammographic screening in a randomized, controlled trial in Sweden. Part I: Background, organization, screening program, tumor findings. In: Day NE, Miller AB, Eds. **Screening for Breast Cancer.** Toronto: Hans Huber; 1988.
21. Tabár L, Faberberg G, Day NE, Holmberg L. What is the optimum interval between mammographic screening examinations?—An analysis based on the latest results of the Swedish Two-County Breast Cancer Screening Trial. **Br J Cancer** 1987; 55: 547–551.
22. Baker LH. Breast cancer detection demonstration project: five-year summary report. **Cancer** 1982; 32: 4–35.
23. Baines CJ, Miller AB, Wall C *et al.* Sensitivity and specificity of first screen mammography in the Canadian National Breast Screening Study: a preliminary report from five centers. **Radiology** 1986; 160: 295–298.
24. Andersson I. Breast cancer screening in Malmö. **Rec Res Cancer Res** 1984; 90: 114–116.
25. Anderson TJ, Lamb J, Alexander F, Lutz W, Chetty U, Forrest APM *et al.* Comparative pathology of prevalent and incident cancers detected by breast screening. **Lancet** 1986; I: 519–523.
26. Holmberg L, Adami H-O, Persson I, Lundström T, Tabár L. Demands on surgical inpatient services after mass mammographic screening. **Br Med J** 1986; 293: 779–782.
27. Aron JL, Prorok PC. An analysis of the mortality effect in a breast cancer screening study. **Int J Epidemiol** 1986; 15: 36–43.
28. Thier SO. Breast cancer screening: a view from outside the controversy. **N Engl J Med** 1977; 297: 1063–1065.
29. Tabár L, Fagerberg CJG, Gad A *et al.* Reduction in mortality from breast cancer after mass screening with mammography. **Lancet** 1985; I: 829–832.
30. Shapiro S. Evidence on screening for breast cancer from a randomized trial. **Cancer** 1977; 39: 2772–2782.
31. Hill D, White V, Jolley D, Mapperson K. Self examination of the breast: is it beneficial? Meta-analysis of studies investigating breast self examination and extent of disease in patients with breast cancer. **Br Med J** 1988; 297: 271–275.
32. Feig SA. Decreased breast cancer mortality through mammographic screening: results of clinical trials. **Radiology** 1988; 167: 659–665.
33. Reidy J, Hoskins O. Controversy over mammography screening. **Br Med J** 1988; 297: 932–933.
34. McLellan GL. Screening and early diagnosis of breast cancer. **J Fam Pract** 1988; 26: 561–568.
35. Bauer M, Schulz-Wendtland R, v.Fournier D. Mammographie-Reihenuntersuchung—Eine Weg zur Reduzierung der Mortalität des Mamma-Karzinoms? **Radiologe** 1988; 28: 95–102.
36. The European Group for Breast Cancer Screening. Guidelines for breast cancer screening. **Clin Radiol** 1987; 38: 217.
37. Tabár L, Fagerberg CJG, Day NE. The results of perodic one-view mammographic screening in a randomized, controlled trial in Sweden. Part II. Evaluation of the results. In: Day NE, Miller AB, Eds. **Screening for Breast Cancer.** Toronto: Hans Huber; 1988.
38. Day NE, Baines CJ, Chamberlain J *et al.* UICC Project on screening for cancer: report of the Workshop on Screening for Breast Cancer. **Int J Cancer** 1986; 38: 303–308.
39. Maguire GP. Possible psychiatric complications of screening for breast cancer. **Br J Radiol** 1983; 56: 284.
40. Dean C, Roberts MM, French K, Robinson S. Psychiatric morbidity after screening for breast cancer. **J Epidemiol Commun Health** 1986; 40: 71–75.

41. Maclean U, Sinfield D, Klein S, Hardnen B. Women who decline breast screening. **J Epidemiol Commun Health** 1984; 38: 278–283.
42. Abholz HH. Risiko-Verminderung als präventives medizinisches Konzept. **Argument (Berlin)** 1984; AS119: 57.
43. Shapiro S, Strax P, Venet L. Periodic breast cancer screening in reducing mortality from breast cancer. **JAMA** 1971; 215: 1777–1785.
44. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten to fourteen year affect of screening on breast cancer mortality. **J Natl Cancer Inst** 1982; 69: 349–355.
45. Wertheimer MD, Constanza ME, Dodson TF, D'Orsi C, Pastides H, Zapka JG. Increasing the effort toward breast cancer detection. **JAMA** 1986; 255: 1311–1315.
46. Wright CJ. Breast cancer screening: a different look at the evidence. **Surgery** 1986; 100: 594–597.
47. Eddy DM, Hasselblad V, McGivney W, Hendee W. The value of mammography screening in women under age 50 years. **JAMA** 1988; 259: 1512–1519.

J Clin Epidemiol Vol. 43, No. 3, pp. 227–234, 1990
Printed in Great Britain.

0895-4356/90 $3.00 + 0.00

A DISSENT FROM DR SCHMIDT'S APPRAISAL OF EVIDENCE ON BREAST CANCER SCREENING

SAM SHAPIRO*

Health Services Research and Development Center, The Johns Hopkins University, Baltimore, MD 21205, U.S.A.

This commentary on Dr Schmidt's paper, "The epidemiology of mass breast cancer screening—a plea for a valid measure of benefit" [1], was stimulated by my reaction on reviewing the paper when it was submitted for publication. The extensiveness of his analysis of benefits and risks in breast cancer screening called for a detailed assessment of his arguments rather than a dissenting note to the editors, a position with which they agreed.

Dr Schmidt probes into published results of the two randomized controlled trials on breast cancer screening in Sweden and draws peripherally on some of the observations in the Health Insurance Plan project to conclude that "even with the assumption that screening can save lives, the net health effect of mass breast screening is questionable and appears to be rather detrimental." The Swedish trials are the two-county study in Kopparberg-Östergötland [2] and the study in Malmö [3], both of which started in the late 1970s, 13–14 years after the HIP trial began.

Dr Schmidt's negative conclusion is based on the observed "excess" in breast cancers detected among study groups of women in the Swedish trials and the additional surgery that results from screening; the small gain when measures of absolute differences between study and control groups in breast cancer mortality are considered instead of relative risk; questionable balance between benefits and risks and costs; increased anxiety among women due to screening; and a variety of other issues.

Some of the points made by Dr Schmidt relate to the efficiency and quality of screening and are the object of attention in the U.S. and other countries where screening is being introduced. Other arguments are a mixture of judgments and interpretations with which there is ample reason to disagree. References are made in the discussion that follows to the HIP study because this trial offers important evidence on the subject, despite its age and mammography's increased capability to detect early breast cancer.

EXCESS BREAST CANCERS AND EXCESS SURGERY DUE TO SCREENING

These are raised by Dr Schmidt as significant issues. In the HIP study [4, 5], they became non-issues since (a) the study and control groups equalized in the numbers of breast cancer between 5 and 7 years after entry (i.e. 1.5–3.5 years after screening ended) and (b) the positive biopsy ratio among women who had a biopsy recommendation on screening (1 confirmed cancer to 4 benign cases) was similar to the ratio in usual clinical practice at that time.

In the Swedish trials, the issue of "excess" breast cancers may never be unequivocally resolved because of the national policy adopted to make screening with mammography generally available, including to the control groups of women. However, there are indications that the problem is not as significant as Dr Schmidt

*Professor Emeritus, Department of Health Policy and Management; Former Director, Health Services Research and Development Center; The Johns Hopkins University School of Hygiene and Public Health.

concludes. Let us take the case of the two-county trial. Cumulative rates of breast cancers detected (invasive plus *in situ* ductal) increased by 1.7 per 1000 women in the study group and 2.6 for the control group during 1985 [6, 7]. The latter may be showing the early effect of screening in the control group but it is the figure (1.7) for the study group that is of special interest. The rate is subject to considerable chance variation and is affected by the extent to which study group women were screened during 1985. But, its level raises question about any measure of "excess" derived from the differential between study and control groups in their 7 year cumulative rates of invasive and *in situ* ductal cancers (16.7 and 13.5 per 1000 respectively).

Dr Schmidt comments on the possible effect of lead time in accounting for the difference but then dismisses it. However, judging from the high prevalence rate of invasive cancer at first screening (5.56/1000 screened women) compared with rates of 2.18 and 1.91 per 1000 person-years reported for study and control groups respectively [8], lead time was unquestionably an important factor and could explain a large part of the observed "excess".

In the Malmö study, the differential in cumulative breast cancer rates over an 8.8 year period also needs to be qualified. Here, the rates are 27.9 and 21.1 per 1000 study and control group women, respectively. The detection rates at first screening (prevalence) is not reported but it would be most surprising if it were not high. Lead time would be present and the observed "excess" would have decreased if termination of the trial, in 1986, had been followed by a period of no, or low levels, of screening for the controls.

To support the proposition that improved mammography could result in detection of large numbers of cancers relative to what becomes clinically known, Dr Schmidt cites the Danish autopsy study of Nielsen *et al.* [9]. However, the fact that in this small sample of cases (83), 70% of the breast cancers, largely *in situ* cancers, were not clinically known, tells us nothing about the magnitude of excess breast cancers resulting from screening with mammography. There is clearly a major difference between autopsy and screening results. In the two-county study, only 8% of the cases detected in the study group were classified as *in situ* ductal; in the Malmö study, the proportion was 16% (the corresponding figures for the control groups were 3 and 11%). In the autopsy study, about two-thirds of the women with malignancies of the breast were found to have ductal or lobular *in situ* carcinomas.

These comparative proportions strongly suggest that the Danish investigators' extremely detailed studies in search of breast cancers have little bearing on the iatrogenic effect of the screening examinations in the two studies. An "excess" may have occurred but the extent to which this happened remains speculative and this is the important point. As for the effect of future improvements in mammography, this too, is speculative. What we do have are experiences in the application of mammography that is in an advanced stage of development.

On the matter of excess surgery reflected by biopsy positive ratios, in some countries it is likely that these ratios will, under current screening and follow-up conditions, be lower than in usual clinical practice. The argument most often advanced for accepting a lowered biopsy positive ratio is that the benefit in terms of lesser surgery and lowered breast cancer mortality outweighs the discomfort, emotional and physical, and the increased costs resulting from the increased numbers of biopsies per confirmed cancer. The unanswerable question is how much of an increase is acceptable and by whom.

None of this should be interpreted as meaning that the issues raised by Dr Schmidt should be ignored; clearly they need attention. A difficulty is that in this era of concern about malpractice suits, in the U.S. at least, "playing it safe" can lead to low biopsy positive ratios not only under screening conditions but in usual clinical practice.

Two other related issues raised by the author as negative aspects of screening involve (a) the referral of 9.7 women for complete mammography per breast cancer detected in the two-county Swedish trial (Kopparberg County, Table 2 of the paper) which is interpreted as an important burden and (b) the occurrence of false positive histology results as shown in the Breast Cancer Detection Demonstration Project (BCDDP) (more on this later) [10]. These, too, cannot be brushed aside. But, in the absence of information on rates of referral for diagnostic mammography by physicians in general or frequency of false positive histology in usual practice, it is not possible to determine the magnitude of the "excesses" that might be attributed to screening.

MAGNITUDE OF REDUCTION IN BREAST CANCER MORTALITY AND ITS SIGNIFICANCE

Central to the author's negative appraisal of screening's benefits is the use of an absolute reduction in mortality from breast cancer per 100,000 women-years. In the two Swedish trials combined, this turns out to be only 6 deaths per 100,000 women-years (Table 4 in the paper). There are two problems with this measure. First, breast cancer mortality in the Swedish and HIP studies included only women who were diagnosed with breast cancer after the start of the trial (the HIP trial excluded all women with prior breast cancer diagnosed; the Swedish studies did not, but eligibility for counts of deaths was restricted to those with a new diagnosis of breast cancer). Accordingly, the exposed to risk for breast cancer mortality was zero at the start of screening. The number increased each year as cases cumulated, but the result is a low annual average mortality rate, which is what the rate per 100,000 person-years represents.

Second, and more important, the main interest in the assessment of screening's effect is in changes in the risk that a woman will die from breast cancer. This calls for cumulating mortality experience per 1000 women. In the HIP trial, where screening ended 3.5 years after entry, there was an absolute reduction of about 2 per 1000 women from the control group's rate of 7–8 per 1000.

Other types of measures have been applied to assess benefit from screening in the HIP trial, all of which provide a different perspective of gain than a rate of 6 per 100,000 women years. These show (a) reduction in person-years of life lost from breast cancer, of 20 per 1000 women, and (b) reduction in the probability of breast cancer death of 11–13 per 100 women with breast cancer (Aron–Prorok life table model [10] applied to breast cancer cases classified by date of women's entry to study).

With respect to the two-county Swedish study, Fig. 2 in the 1987 paper by Tabar *et al.* [11] indicates that the absolute difference between the cumulative mortality rates for study and control groups of women, at the end of 8 years since randomization, was 88 per 100,000 women.

The decreases cited above may still seem small but in the U.S., replication of the HIP trial on a national scale (baseline plus 3 annual screenings among the 30 million women aged 40–64 years with two-thirds having at least one examination) would have resulted in a decline in breast cancer deaths over a 10 year period from about 193,000 to 147,000, i.e. a decrease of about 46,000. The calculation covers an interval extending long after screening ended and it is reasonable to assume that the decrease would have increased, absolutely and relatively, if screening had continued.

The numbers are not large in comparison with the over 2 million deaths from all causes in a cohort of women 40–64 years of age followed for 10 years. But, the reduction takes on different significance when we consider the fact that in the U.S., the probability over a lifetime of women having a diagnosis of breast cancer is 10% and that mortality from this condition ranks first or second among deaths from cancers of all types in the female population.

Dr Schmidt attempts to place the reduction in breast cancer mortality attributable to screening in another perspective by drawing comparisons with daily risks among construction workers and car drivers. This is obviously spurious; even the eradication of all breast cancer deaths would presumably be unimportant considering other daily risks.

BENEFITS, RISKS, COSTS

Dr Schmidt calls attention to increased mortality from causes other than breast cancer in the study group of women with breast cancer. This is so in the Swedish and HIP trials. But, when we add these deaths to breast cancer deaths, we still find in the HIP study that at the end of 10 years from entry, the study group has 23% fewer deaths among cases detected within 7 years from entry; over the longer term (18 years) when general mortality reaches high levels, given the ages involved, the reduction becomes smaller (15%).

The observed difference after 7 years in total mortality among breast cancer cases in the two-county study is in the same direction as in the HIP study but the OR in Table 5 of the paper, 0.92, indicates that the reduction in breast cancer deaths in the study group was almost balanced by an increase in deaths from other causes. This is too pessimistic a picture. In the two-county study an adjustment is needed in the count of deaths from other causes in the control group to allow for the larger number of women with breast cancer in the study group. The rates per 100 cases in the table are directed at this issue; the rates per 10,000 women in the

first and third columns of the table are not. A lower ratio, 0.86, than in the paper, is obtained by adjusting for differences in numbers at risk for mortality from other causes and, in the case of breast cancer deaths, for the somewhat smaller population in the control group. The ratio would be reduced further in a model that took into account the longer average period of exposure for other cause mortality in the study group due to the earlier detection of breast cancer.

In terms of numbers, the adjusted figures show 53 fewer breast cancer deaths vs 20 additional deaths from other causes among study group women with breast cancer. Why there should be an excess of deaths from other causes is conjectural but the preceding makes Dr Schmidt's summary statement, "there are no data that breast cancer prevents significantly more deaths than it eventually may cause" highly questionable.

In any event, the focus is on reduction in breast cancer mortality which has remained unchanged in the past 40 or more years in the U.S.; in some countries the rate has been increasing. By way of contrast, large decreases in mortality from coronary heart disease and cerebrovascular conditions are occurring and this is being hailed as a major accomplishment. A 20–30% reduction in breast cancer mortality achievable by widespread screening would also be viewed as a highly significant social benefit.

Dr Schmidt raises the question of costs in the context of general statements about cost benefit. No formal analysis is presented but he is critical of cost estimates (as prepared in Sweden) that do not consider absenteeism from work to participate in screening, effects of stress and anxiety related to screening, lengthening of the disease period because of earlier detection, and "excess" cancers and surgery. On the other hand, he does not refer to economic benefits of reduced mortality from breast cancer, lesser surgery, and improved functioning of women with breast cancer detected early.

A full discussion of cost–benefit analysis requires a different forum than the present exchange in views but it should be noted that the subject of costs has not been ignored by any means. Mammography is central to breast cancer screening programs being initiated and this modality is costly, a factor that has received a great deal of attention. In the U.S., the charge per screening mammographic examination has been reduced in many parts of the country to $50, i.e. one-third or less of what it had been; research is underway to determine the effectiveness of setting priorities for screening based on risk factors; and increasingly, HMOs where screening can be organized on a more efficient basis than in the general community are including screening with mammography as a standard benefit.

Cost-effectiveness has been dealt with over a number of years in the U.S. Eddy's first efforts were carried out almost 10 years ago; more recently, in a National Cancer Institute publication [12] and most recently, in the **JAMA** article mentioned by Dr Schmidt [13]. A detailed analysis by the U.S. Office of Technology Assessment preceded the enactment in 1988 of legislation by Congress to reimburse up to $50 per examination for screening mammography every 2 years for Medicare beneficiaries, a large majority of whom are 65 years of age or older [14].

In the U.K., the policy to provide screening with mammography was based on the Forrest report [15] which considers, in great detail, many of the measures in Dr Schmidt's paper and relates economic and other resource requirements to person-years of life saved. The conclusion is that "the estimates for cost per life-year gained for breast cancer screening are not dissimilar to other health service activities currently undertaken." The point is also made that "expansion of certain other services may offer higher returns on resources spent" but the decision was to introduce periodic screening with mammography at ages 50–64 through a carefully planned, incremental program.

INCREASED ANXIETY AMONG WOMEN DUE TO SCREENING

This issue is returned to several times by Dr Schmidt and will be considered here, although many of its facets relate to how we assess benefits and the matter of "excess" surgery which have already been dealt with. Dr Schmidt's arguments add up to a charge that screening unduly creates anxiety among women. However, there is every indication that the level of worry is high even without screening with mammography. Surveys in the U.S. show that a large majority of the women report they are practicing breast self-examination in one form or another and about half had clinical breast examinations during the prior year [16]. In Sweden, increasing proportions of the women in

the control groups were having mammographic examinations before the trials ended [17, 18].

Dr Schmidt cites the finding in a study in Edinburgh [19] that "psychiatric morbidity produced by breast cancer screening among complying women may be considered a minor problem;" only 8% felt that screening had made them feel anxious about developing breast cancer. However, he sets this observation aside with the report that, in an earlier study in Edinburgh, 30% of the women invited for screening did not comply "because of the anxiety a mere invitation can cause." Actually, the 30% refers to the total group of non-attenders interviewed. About 39% of these women were reported to have "expressed fear", varying in content from concern about what would happen if cancer were found to a misinterpretation of the letter inviting them to a screening examination. Many other reasons were given by large proportions of the non-attenders, on the average, 2 major categories were mentioned per respondent, e.g. practical reasons (46%), screening not necessary (38%), negative reactions to physical examinations or other aspects of care (37%). It seems reasonable to conclude that we are not faced with a highly prevalent problem, but that initiation and conduct of screening programs need to be sensitive to the fact that anxieties may surface among some women.

Dr Schmidt also includes as a burden (i.e. worry), additional disease years imposed by earlier detection of breast cancer through screening. Of course, there are more such years than without screening; the argument in favor of screening is that the "burden" is outweighed by lowered mortality from breast cancer and the additional years of life gained following lesser primary and adjuvant therapy.

Study design

It is not clear why Dr Schmidt asserts that the two-county trial in Sweden was faulty and did not adhere to a pre-defined protocol. A randomized controlled trial (RCT) can be based on allocations of geographic subareas; the individual need not be the primary sampling unit. The published papers on this trial do not appear to support Dr Schmidt's conclusion that the results related to "changing sub-populations."

The case–control design of studies in The Netherlands [21, 22] and Italy [23] is characterized as uninformative on the extent of the effect of screening because of uncontrolled biases. Research is, indeed, needed on the relationship between odds ratios derived from case–control studies and those obtained in RCTs but at a minimum the data now available from the former on efficacy of breast cancer screening can be taken as reinforcing the results of the RCTs. Dr Schmidt does not refer to "reinforcement" but he does agree that "case–control studies may be able to demonstrate a true breast cancer mortality reduction."

More on burden and risk of breast cancer screening

Dr Schmidt uses the experience in Kopparberg to illustrate the extensive effort required to achieve a reduction in breast cancer mortality; he states "7000 screening mammographies were needed for the promotion of a reduction in one breast cancer death." This ratio is probably 2–3 times the actual figure; approximately 35,000 women participated in each of the first three rounds of screening examinations, some women also had their fourth screening [24], and there was a reduction of about 37 breast cancer deaths in the study group (after adjustment for lower sample size of the control group). In the HIP study, the ratio was 1500–2000 screenings to one breast cancer death avoided.

The problem of false positive histology among cases initially classified as "minimal" breast cancer cases is raised by Dr Schmidt based on results of the review of pathology tissue slides in the Breast Cancer Demonstration Project [25]. Dr Schmidt includes in his figure on false positives (17%), cases classified as "borderline" because of failure to resolve differences of opinion in the panel of pathologists; also, he does not consider the results of an additional review of the panel. At the end of the review process, about 9% of those designated in the project as "minimal" cancers were not classified as cancer. The reduction from 17 to 9% is important only in that it reflects a tendency for Dr Schmidt to overstate his case against screening. There are difficulties in diagnosing small lesions and the reviewers recommended that "concurrent or multiple pathologic opinions should be obtained in borderline lesions, possibly cancer. . . ." This is believed to have contributed to widespread use of two-stage breast surgery.

Another indicator of Dr Schmidt's negative tendency is found in his discussion of radiation hazard of modern mammography. He refers to this as "very small, causing $\leqslant 1\%$ of breast cancers)." But, one of the references in Dr

Schmidt's paper [26] provides an estimate of 1–2 per 1000 breast cancers in a program utilizing two-view, low-dose film mammography on an annual basis after age 40. The 1% figure is applicable to xeroradiography which has not accounted for more than a small proportion of mammographies performed and is falling into disuse.

The Malmö trial

This RCT is the only study on the efficacy of breast cancer screening that has shown no benefit in the study group. Women were 45–69 years of age at entry; follow-up averaged 8.8 years; at the end of the trial, breast cancer mortality in the study and control was the same; previously, the rate was greater in the study group [27]. More time will be needed to determine whether the higher rates of stages II–IV breast cancers in the control group vs the study group, starting several years after the trial began, presage differentials in breast cancer mortality in favor of the study group. A favorable differential appears to have started already among women aged 55 or older at entry but not for the women 45–54.

The authors of the paper on the Malmö RCT state that "our data support previous studies showing that invitations to mammographic screening for breast cancer may lead to reduced mortalities from breast cancer, at least in women aged 55 and over." In the HIP trial, the study group of women, 40–49 years of age at entry, started to show an advantage substantially later than those 50–59 and differences exist about the strength of this evidence alone to support screening at these ages. However, one view is that longer periods of follow-up are needed for younger than older women because of possible differences in the natural history of breast cancer. Resolution of this question would require additional years of follow-up in the two-county and Malmö studies in Sweden, perhaps for the same duration as in the HIP RCT, 18 years from entry. Also, one component of the Breast Screening Study in Canada is designed to test the efficacy of screening women aged 40–49 [28].

CONCLUDING COMMENTS

Dr Schmidt assesses the results on breast cancer screening from the two-county and Malmö randomized controlled trials in Sweden and draws a highly negative picture of the value of screening. His arguments are based on two considerations:

(a) benefit is small when based on a "valid measure of benefit," defined as the absolute reduction in breast cancer deaths per unit population.
Comment. Breast cancer's low annual incidence rate, about 2 per 1000 adult women makes it certain that an annual absolute measure of change would show any decrease in cancer mortality to be small. However, it is widely recognized that breast cancer is a serious health problem that requires high priority for efforts to reduce the associated mortality. In this context, relative change in breast cancer deaths that can be credited to screening is a key measure, and various studies show a 20–30% potential for reduction in such deaths.

(b) risks in screening, principally, excess in cases diagnosed and in surgery, increased mortality from other causes among women with breast cancer, added disease years due to earlier detection of breast cancer, increased anxiety, and the burden of investigation (i.e. large numbers of women screened and referred for additional testing and treatment) reach high levels and outweigh the benefits from screening.
Comment. The significance of the risks identified by Dr Schmidt is questionable for the screening studies completed or underway. This conclusion is based on a reexamination of available data. Excess in breast cancers diagnosed and treated cannot be quantified at this point, in the Swedish studies; contrary to Dr Schmidt's assessment, mortality from other causes among study group women with breast cancer is substantially smaller than the reduction in breast cancer; added disease years become relevant only if screening's benefit is small; increased anxiety may occur among some women but the meager evidence available suggests it is not a large problem.

With respect to "burden of investigation", we have to bear in mind that the purpose of screening is "to sort out apparently well persons who probably have a disease from those who probably

do not" [29], and the burden will be high in general population screening for low prevalence and incidence conditions such as breast cancer. In the end, the critical question is whether the benefit is worth the effort, a judgment that is affected not only by costs and risks but by the degree of certainty that a gain can be achieved and how we value the benefit demonstrated. In an increasing number of countries, agreement is being reached after extensive deliberation, that the balance is in favor of screening.

In conclusion, Dr Schmidt's arguments do not provide a sound basis for reversing the actions being taken to have large numbers of women receive breast cancer screening examinations that include mammography. Programs and policies differ in the ages targeted, periodicity of screening, type of mammography used, organization, financing, and responsibility for screening.

The decisions are not immutable. The content of screening programs may change as new results become available from studies still in progress. Also, of great importance will be observations on the quality, effectiveness, and costs of breast cancer screening as it moves from the experimental stage to practice in the community. In this connection, first steps have been taken to develop an international uniform data base that will provide information to assess screening experiences [30]. All of this means that while more will be learned about screening, particularly after it reaches the general population, we know enough to move ahead with the diffusion of screening.

REFERENCES

1. Schmidt JG. The epidemiology of mass breast cancer screening—a plea for a valid measure of benefit. **J Clin Epidemiol** 1990; 43: 215–225.
2. Fagerberg CJG, Tabar L. The results of periodic one-view mammography screening in a randomized controlled trial in Sweden. Part 1: Background, organization, screening program, tumor findings. Tabar L, Fagerberg CJG, Day NE. Part 2: Evaluation of the results. In: Day NE, Miller AB, Eds. **Screening for Breast Cancer.** Toronto: Hans Huber; 1988.
3. Andersson I, Aspergren K, Janzon L *et al.* Mammographic screening and mortality from breast cancer: The Malmö Mammographic Screening Trial. **Br Med J** 1988; 297: 943–948.
4. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selection, follow-up and analysis in the Health Insurance Plan Study: a randomized trial with breast cancer screening. In: Garfinkel L, Ochs O, Mushinski M, Eds. **Selection, Follow-up, and Analysis in Prospective Studies: A Workshop**. NIH Publication 85-2713; National Cancer Institute Monograph 67. Washington, D.C.: DHHS, PHS; 1985.
5. Shapiro S, Venet W, Strax P, Venet L. **Periodic Screening for Breast Cancer: The Health Insurance Plan Project and Its Sequelae, 1963–1986.** Baltimore, Md: The Johns Hopkins University Press; 1988.
6. Tabar L, Fagerberg G, Gad A *et al.* Reduction in mortality from breast cancer after mass screening with mammography. **Lancet** 1985; 1: 829–832.
7. Tabar L, Fagerberg G, Day N, Holmberg L. What is the optimum interval between mammographic screening examinations?—An analysis based on the latest results of the Swedish Two-county Breast Cancer Screening Trial. **Br J Cancer** 1987; 55: 547–551.
8. *Op. cit.* Ref. [2].
9. Nielsen M, Jensen J, Andersen J. Precancerous and cancerous breast lesions during lifetime and at autopsy. **Cancer** 1985; 54: 612–615.
10. Aron J, Prorok P. An analysis of the mortality effect in a breast cancer screening study. **Int J Epidemiol** 1986; 15(1): 36–41.
11. *Op. cit.* Ref. [7].
12. Eddy DM. A computer-based model for designing cancer control strategies. In: Greenwald P, Sondik E, Eds. **Cancer Control: Objectives for the Nationa, 1985–2000.** National Cancer Institute Monograph 2. Washington, D.C.: DHHS, PHS; 1986.
13. Eddy DM, Hasselblad V, McGivney W, Hendee W. The value of mammography screening in women under age 50 years. **JAMA** 1988; 259(10): 1512.
14. U.S. Congress Office of Technology Assessment. **Breast Cancer Screening for Medicare Beneficiaries: Effectiveness, Costs to Medicare and Medical Resources Required.** November 1987.
15. Forrest P. **Breast Cancer Screening: Report to the Health Ministers of England, Wales, Scotland, and Northern Ireland.** London: Her Majesty's Stationery Officer; 1987.
16. Thornberry OT, Wilson RW, Golden PM. Health promotion data for the 1990 objectives: estimates from the National Health Interview Survey of health promotion and disease prevention: U.S., 1985. **NCHS Advance Data,** Vital and Health Statistics of NCHS, USDHHS, PHS; 1986: No. 126.
17. *Op. cit.* Ref. [6].
18. *Op. cit.* Ref. [3].
19. Dean C, Roberts MM, French K, Robinson S. Psychiatric morbidity after screening for breast cancer. **J Epidemiol Commun Health** 1986; 40: 71–75.
20. Maclean U, Sinfield D, Klein S, Harden B. Women who decline breast screening. **J Epidemiol Commun Health** 1984; 38: 278–283.
21. Collette H, Rombach J, Day N, deWaard F. Evaluation of screening for breast cancer in a non-randomized study (the DOM Project) by means of a case–control study. **Lancet** 1984; 1: 1224–1226.
22. Verbeek A, Hendriks F, Holland R, Strumans J, Mravanuac M, Day N. Reduction of breast cancer mortality through mass screening with modern mammography: first results of the Nijmegen Project, 1975–1981. **Lancet** 1984; 1: 1222–1224.
23. Palli D, delTurco MR, Buetti E, Carli S, Cialto S *et al.* A case–control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). **Int J Cancer** 1986; 38: 501–504.
24. *Op. cit.* Ref. [2].
25. Beahrs O, Shapiro S, Smart C. Report of the Working Group to Review NCI/ACS Breast Cancer Detection Demonstration Project. **J Natl Cancer Inst** 1979; 62: 673–675, 702–704.

26. Gohagan JK, Darby WP, Spitznagel EL, Monsees BS, Tome AE. Radiogenic breast cancer effects of mammographic screening. **J Nat Cancer Inst** 1986; 77(1): 71–76.
27. *Op. Cit.* Ref. [3].
28. Miller AB, Howe G, Wall C. The national study of breast cancer screening. **Clin Invest Med** 1981; 4: 227–258.
29. Wilson JMG, Jungner G. **Principles and Practice of Screening for Disease.** Public Health Papers 34. Geneva: WHO; 1968.
30. **International Workshop on Information Systems in Breast Cancer Detection.** Sponsored by the Center for Devices and Radiological Health, Food and Drug Administration and the Division of Cancer Prevention and Control, National Cancer Institute, U.S. Department of Health and Human Services, Rockville, Md, 8–10 December 1988.

J Clin Epidemiol Vol. 43, No. 3, pp. 235–239, 1990
Printed in Great Britain.

0895-4356/90 $3.00 + 0.00

Response

RESPONSE TO DR SHAPIRO'S DISSENT

JOHANNES G. SCHMIDT

Center for Clinical Epidemiology, University of Newcastle, NSW 2308, Australia and
Institut Universitaire de Médecine Sociale et Préventive, CH-1005 Lausanne, Switzerland

It may be helpful to begin my response to Dr Shapiro's dissent [1] with a brief outline of the purpose of "clinical epidemiology". Traditionally, clinical medicine focuses on mechanisms of diseases in individuals. This perspective, quite often, tends to ignore the epidemiological and quantitative significance of such mechanisms. Therefore, it is often assumed that early detection necessarily interrupts the mechanisms of further dissemination and that early detection necessarily represents a benefit. Classical epidemiology, on the other hand has mainly focused on a population perspective and, therefore, welcomes as a key success the statistically significant reduction in mortality of a disease that is a serious health problem. This perspective tends to ignore the clinical significance of such a success and is unaware of the individual perspective and the total health impact on the potential patients. Both traditional approaches, thus, tend to see breast cancer screening from a limited and biased perspective. It is the relatively innovative approach of "clinical epidemiology" which brings the various perspectives together, putting the relevant clinical events both into a quantitative and individual perspective. However, this change of paradigm, though inevitable, takes time and is facing considerable dissent.

In his dissent Dr Shapiro maintains the perspective of classical epidemiology. He makes a plea that "the focus is on reduction in breast cancer mortality which has remained unchanged in the past 40 or more years in the U.S." [1]. However, many would agree that the prime focus should be on the total health effects of any intervention. The crucial question, therefore, is whether screening as a method to reduce breast cancer mortality does truly more good than harm. This question, although more complex and complicated, is what we have to discuss and resolve.

In an attempt to address this latter question, my paper [2] provides an enumeration of beneficial and detrimental effects of breast cancer screening, As such, it is based entirely on real observation from randomized controlled trials, for a total active trial period of up to 9 years. In order to achieve the highest statistical confidence it uses a pooled analysis of the adequate randomized trials (as to the HIP trial, see below). Sensitivity, specificity, and predictive value of the screening test in question, i.e. of modern mammography, are calculated according to the rather favourable screening performance achieved in the Swedish trials. In fact, other screening trials showed less favourable results in this respect [2, 3]. Generally, the "burden of investigation" may be rather greater than estimated from the two Swedish trials.

MAGNITUDE OF MORTALITY REDUCTION: RELATIVE AND ABSOLUTE RISKS

The reference made to the prevention of coronary heart disease and cerebrovascular conditions to support the case of breast cancer screening [1] is a good example to illustrate the fallacy of insisting on the use of relative risks as a "key measure". The proportionate mortality

*All correspondence should be addressed to: J. Schmidt, Gesundheitsplan HMO (Swiss HMO Health Plan), Group Practice Schmiedhof Zweierstrasse 136/138, CH-8003 Zurich, Switzerland.

for coronary heart disease and cerebrovascular disease in the groups at which preventive action is targeted is regularly in the order of 50–70% (56% in hypercholesterolemic males of the LRC-CPPT trial [4]; 56% in mild to moderate hypertensives of the MRC trial [5], and 68% of the Australian trial [6]; 57% in mild hypertensives of the HDFP programme [7]; 69% in elderly hypertensives [8]; and 68% in hypertensive males of the Oslo study [9]). This compares to 3% proportionate breast cancer mortality in the Swedish trials (2.8% in the two counties; 3.6% in Malmö). A 20% reduction in cardiovascular mortality would correspond to a 10–14% reduction in total mortality, whereas a 20% breast cancer mortality reduction represents little more than 0.5% reduction in total mortality. The 20% reduction in terms of relative risks, therefore, means quite a different thing in both cases. This shows, indeed, how misleading relative risks can be.

Although Dr Shapiro mentions the cumulative mortality and the count of life years gained, he does not discuss the relevance of these measures. My failure to address the "cumulating mortality experience per 1000 women" is considered to be the "more important" problem [1]. Actually, the latter is also an absolute risk measure. It refers to a lifetime perspective and is certainly a measure of some relevance since it provides essentially the same information as the mortality experience per a certain number of women-years (simple absolute risk). However, it is founded on uncertain mathematical extrapolation and it also fails to take into account the time–distance to the event and associated discounting (see below). Simple absolute risks are easy to obtain and they allow a direct comparison of the number of beneficial and adverse effects.

Moreover, life table models tend to exaggerate the individual chance of experiencing an event [10]. The 10% lifetime probability of having breast cancer, mentioned by Dr Shapiro [1], is a misleading figure because it represents the probability only of the minority of women who get very old. Further, as screening does not reduce but rather increases the incidence of overt breast cancer, only the mortality figure can inform us about the potential significance of screening for this health problem. 3.4% of all women in Sweden, and not 10%, die from breast cancer (4.1% in the U.S.) [11]. It should not be overlooked that less than half of the women with breast cancer eventually die from their disease [12]. Moreover, approximately half of all breast cancers are not amenable to screening which means that barely 2% of female mortality can be targeted by breast cancer screening, and only approx. 20% of it can be prevented (i.e. *ca* 1 out of 250 deaths). Telling women that they have a 10% risk of developing breast cancer in order to motivate them to comply with screening recommendations is a bit misleading, in individual and public health terms.

Dr Shapiro considers the use of absolute risks as a problem because screening studies had a "low annual average mortality rate" [1]. However, if one is to analyse the effects of screening, it is evident that only new breast cancer cases are of relevance because screening cannot have an effect on already existing breast cancer cases. The exclusion of women with already existing breast cancer does not appreciably reduce the denominator, but the numerator is reduced. Therefore, mass screening will necessarily be limited to women with a low average mortality rate.

The mention of 88 prevented deaths per 100,000 women [1], instead of the 6 per 100,000 women years, may cause confusion. This figure, however, is in accordance with the figure in my paper since the 88 prevented deaths refer to a period of 8 years, thus 800,000 women years. In addition, Dr Shapiro mentions only the trial with the more favourable results. Dr Shapiro also claims that I overestimated the number of mammographies per breast cancer death prevented more than two-fold [1]. However, he singles out again the sub-study with the most favourable result (i.e. Kopparberg). The Malmö Study showed virtually no mortality reduction [13], and the number of deaths prevented in Östergötland was less than half the number in Kopparberg [14].

EXCESS BREAST CANCERS

The breast cancer excess in the screened groups is neither hypothetical nor speculative, but demonstrable in all trials using sensitive modern mammography (24% increase at 7 years in the two counties [15], 32% at 9 years in Malmö [13], 44% at 5 years in Stockholm [3], 51% at 7 years in the non-randomized U.K. trial [16]). The question as to what this excess means in the end, practically and theoretically, remains difficult to resolve.

The HIP study was the first randomized trial to show a mortality reduction through breast cancer screening, an observation that was confirmed by the Swedish trials (combined Mantel–Haenszel OR of all three trials: 0.75, $p = 0.004$). The HIP trials, however, lacks external validity for the question of the investigation load of modern screening, including the problem of possible excess breast cancers. The screening sensitivity and the cancer detection rate in this trial [17] were dramatically and significantly lower than in Sweden [18] ("intention-to-screen" sensitivity: 44 vs 67%, $p \ll 0.001$; detection rate per 1000 invited women in prevalence screen: 1.7 vs 4.6, $p \ll 0.001$). Therefore, the fact that a possible cancer excess "became a non-issue" [1] in the HIP study has little relevance for present-day breast screening.

Moreover, the division of screening into prevalence screening (i.e. initial screen) and incidence screening (i.e. subsequent screens) cannot be assumed to be a valid method of assessing the lead time effect. This method, although the conventional approach, is based on the assumption that possible overdiagnosis is non-existent during the intitial screen. However, the observation of considerable false cancer diagnosis, as well as the observation that as many as 70% of breast cancers detected at autopsy in one study were clinically silent, both, emphasize that overdiagnosis is possible, be it at incidence or at prevalence screening. There is no need to debate the extent of false histology, because my paper used this observation only to draw attention to the potential of overdiagnosis and not to speculate about the extent of this effect. However, it should be noted that equivocal "borderline" histology, of course, may also contribute to false cancer diagnosis and associated overdiagnosis. Dr Shapiro's complete omission of borderline histology as a source of overdiagnosis is not quite correct.

It is often argued that potential gains far in the future must be discounted. The significance of breast cancer deaths potentially prevented in the distant future will further diminish because of competing mortality from other causes. For example, breast cancer deaths in Sweden constitute 12% of the total mortality in the 45–54 age group, 9% in the 55–64 age group, 5% in the 65–74 group, and only 2% in women of age 75 and above [11]. The Malmö trial [13] shows this kind of "epidemiological discounting" from yet another perspective: the earlier the stage and the better the prognosis the higher is the proportion of deaths from other causes (stage III: 19%, stage II: 30%, stage I: 72%). In fact, other authors analysing the HIP study have observed that "the number of cases in the two groups (i.e. study and control) may never equalize because other causes of death effectively censor incidence from breast cancer" [19].

What we have is the experience of a 9-year perspective in which the occurrence of an excess cancer is almost ten times as likely as the occurrence of a prevented breast cancer death [2].

THE PROBLEM OF ADVERSE MORTALITY EFFECTS

The possible occurrence of small increases in mortality attributable to screening becomes relevant only because the breast cancer mortality reduction itself is very small. Data have been offered to suggest that adverse mortality effects are a possibility [2], but the significance and the extent of such an effect remains speculative. However, the screening trials are too small to measure such effects. For example, if a 33% breast cancer mortality reduction had occurred in the Malmö trial this would account for 22 prevented deaths. If screening at the same time would cause an increase in cardiovascular mortality of only 3.3%, this would make up for 22 deaths as well. The 33% breast cancer mortality reduction would be significant ($p = 0.036$), but a cardiovascular mortality increase responsible for the same number of deaths would be statistically highly insignificant ($p = 0.55$). In fact, there were 48 more deaths from cardiovascular disease in the screened group [13], but even this mortality increase is statistically not significant ($p = 0.16$).

A very large study of 424,127 deceased cancer patients was able to measure a statistically significant effect of cancer diagnosis on suicide mortality, notably in Sweden [20]. In women with a diagnosis of breast cancer, the suicide rate was increased 1.4-fold (95% confidence interval: 1.1–1.7) while the strongest increase was found in the first years after diagnosis, 15.5-fold in the first year and 7.0-fold in the second and third year since diagnosis (latter figures pertain to all cancer). This would mean that there is some increase of suicide in the screening group associated with the increase of diagnosed breast cancer. In absolute numbers, this increase in suicide is very small, but it

proves that screening may indeed have adverse mortality effects and that the net mortality gain may be further diminished. This has regularly been ignored in earlier appraisals of breast cancer screening. The extent of such adverse mortality is unknown because we are unable to test this question. The problem, however, cannot be brushed aside.

"INTANGIBLE COSTS"

Breast disease, whether malignant or not, is associated with a lowered self esteem, poorer body image and impaired sexual relationships [21]. A suspicious mammography will have an impact on women's quality of life both with and without an eventual diagnosis of breast cancer. Of course, screening does not only increase the amount of overt breast cancer but also the amount of overt benign breast disease, the latter probably even to a greater extent: In Kopparberg [18], Östergötland [22] and Stockholm [23], there were 3.3–3.5 benign lumps found per breast cancer. The quality of life may be only slightly impaired by the diagnosis of a benign breast disease. However, these life years with only slight impairment will be much more numerous than the few cases in which screening offers an improvement in the quality of life. Increased anxiety is common both in compliers of a screening programme and in non-compliant women who have merely been invited. Initially, about 20% of all women may suffer from this "minor" problem [24, 25]. Our zeal to screen can prevent breast cancer death, but in the 10 year perspective it will cause such minor health impairment 300 times more often. Dr Shapiro's conclusion that such minor health problems are not highly prevelant [1] overlooks again the crucial question of the prevalence of this problem relative to the prevalence of benefit.

In terms of "costs", my paper almost exclusively deals with the issue of health costs or "epidemiological costs". The over-optimistic results of many formal cost–benefit appraisals can be explained by the fact that time costs were omitted [26] or that a guess was made of the "intangible costs" by assuming that they were outweighed by the mortality gain or by the benefit of reassurance [27]. In fact, this latter benefit does not even exist [2].

"Intangible costs" have been insufficiently investigated to allow a conclusive answer to the crucial question of whether the net burden of suffering can be reduced by screening. This critical question has not been resolved by empirical investigation, but by beliefs, guess work, or even wishful thinking, mostly by parties with vested interests (including the interest of the "helping profession" to promise help). It is unknown how (individual) women value the benefit of a prevented breast cancer death (in 10 years, this will occur with a chance of less than 1 per 1000) if they could compare it to the "burden" to achieve this benefit (e.g. in 10 years, a positive screening mammography will occur with a chance of about 150 per 1000, an overt cancer that would not have been detected without screening will occur with a chance of 5 per 1000, to mention but a few [2]).

I believe that many women would choose against screening if they were properly informed of the true risks and benefit of screening in absolute risk terms.

The net health effect of breast cancer screening, indeed, remains questionable, and when scrutinized carefully appears to be even detrimental. The dissent by Dr Shapiro, although a valuable contribution to the debate, provides no convincing arguments to show a clear benefit of mass breast screening. More confidence could be achieved by unbiased research about gain and loss of quality adjusted life years.

REFERENCES

1. Shapiro S. A dissent from Dr Schmidt's appraisal of evidence on breast cancer screening. **J Clin Epidemiol** 1990: 43; 227–234.
2. Schmidt JG. The epidemiology of mass breast cancer screening: A plea for a valid measure of benefit. **J Clin Epidemiol** 1990: 43; 215–225.
3. Frisell J, Eklund G, Hellström L, Glas U, Somell A. The Stockholm Breast Cancer Screening Trial: 5-year results and stage at discovery. **Breast Cancer Res Treat** 1989; 13: 79–87.
4. The Lipid Research Clinics Coronary Primary Prevention Trial results. I. Reduction in incidence of coronary heart disease. **JAMA** 1984; 251: 351–364.
5. Medical Research Council Working Party on Mild to Moderate Hypertension. MRC trial of treatment of mild hypertension: principal results. **Br Med J** 1985; 291: 97–104.
6. Management Committee. The Australian therapeutic trial in mild hypertension. **Lancet** 1980; 1: 1261–1267.
7. Hypertension Detection and Follow-Up Program Cooperative Group. Five-year findings of the hypertension detection and follow-up program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension. **JAMA** 1979; 242: 2562–2571.
8. Amery A, Birkenhager W, Brixko P, Bulpitt C, Clement D, Deruyttere M *et al.* Mortality and morbidity results from the European Working Party on High Blood Pressure in the Elderly trial. **Lancet** 1985; 1: 1349–1345.

9. Helgeland A. Treatment of mild hypertension: a five year controlled drug trial. The Oslo study. **Am J Med** 1980; 69: 725–732.
10. Love RR. The risk of breast cancer in American women (letter). **JAMA** 1987; 257: 1470.
11. WHO. **World Health Statistics 1987.** Geneva: WHO; 1987.
12. Lundgren B. Diagnosis and screening in breast cancer; a review. **Eur J Cancer Clin Oncl** 1983; 19: 1709–1710.
13. Andersson I, Aspegren K, Janzon L *et al.* Mammographic screening and mortality from breast cancer: the Malmö Mammographic Screening Trial. **Br Med J** 1988; 297: 943–948.
14. Tabár L, Fagerberg CJG, Day NE. The results of periodic one-view mammographic screening in a randomized, controlled trial in Sweden. Part II. Evaluation of the results. In: Day NE, Miller AB, Eds. **Screening for Breast Cancer.** Toronto: Hans Huber; 1988.
15. Fagerberg CJG, Tabár L. The results of periodic one-view mammographic screening in a randomized, controlled trial in Sweden. Part I. Background, organization, screening program, tumor findings. Day NE, Miller AB, Eds. **Screening for Breast Cancer.** Toronto: Hans Huber; 1988.
16. U.K. Trial of Early Detection of Breast Cancer Group. First results on mortality reduction in the U.K. trial of early detection of breast cancer. **Lancet** 1988 II: 411–416.
17. Shapiro S. Evidence on screening for breast cancer from a randomized trial. **Cancer** 1977; 39: 2772–2782.
18. Tabár L, Gad A, Holmberg L, Ljungquist U. Significant reduction in advanced breast cancer—results of the first seven years of mammography screening in Kopparberg, Sweden. **Diagn Imag Clin Med** 1985; 54: 158–164.
19. Aron JL, Prorok PC. An analysis of the mortality effect in a breast cancer screening study. **Int J Epidemiol** 1986; 15: 36–43.
20. Allebeck P, Bolund C, Ringbäck G. Increased suicide rate in cancer patients. **J Clin Epidemiol** 1989; 42: 611–616.
21. Morris T, Steven-Greer H, White P. Psychological adjustment to mastectomy (a 2 year follow up study). **Cancer** 1977; 40: 2381–2387.
22. Fagerberg G, Baldetorp L, Gröntoft O, Lundström B, Manson JC, Nordenskjöld B. Effects of repeated mammographic screening in breast cancer stage distribution. **Acta Radiol** 1985; 24: 465–473.
23. Frisell J, Glas U, Hellström L, Somell A. Randomized mammographic screening for breast cancer in Stockholm. **Breast Cancer Res Treat** 1986; 8: 45–54.
24. Dean C, Roberts MM, French K, Robinson S. Psychiatric Morbidity after Screening for Breast Cancer. **J Epidemiol Commun Health** 1986; 40: 71–75.
25. Maclean U, Sinfield D, Klein S, Hardnen B. Women who decline breast screening. **J Epidemiol Commun Health** 1984; 38: 278–283.
26. Day NE, Baines CJ, Chamberlain J *et al.* UICC project on screening for cancer: report of the Workshop on Screening for Breast Cancer. **Int J Cancer** 1986; 38; 303–308.
27. Forrest P. **Breast Cancer Screening.** Report to the Health Ministers of England, Wales, Scotland, and Northern Ireland. London: Her Majesty's Stationery Office; 1986.