Age-Specific Reduction in Breast Cancer Mortality by Screening: An Analysis of the Results of the Health Insurance Plan of Greater New York Study[1,2]

J. D. F. Habbema,[3] Gerrit J. van Oortmarssen,[3] Dick J. van Putten,[4] Jacobus T. Lubbe,[3] and Paul J. van der Maas[3,5]

ABSTRACT—The 14-year follow-up mortality results from the randomized breast cancer screening trial of the Health Insurance Plan of Greater New York (HIP) were analyzed with respect to the problem of age-specific screening effects. Mortality reduction was measured in three different ways and appears to be homogeneous across age groups. This finding challenges the widespread opinion that the results of the HIP study support the conclusion that breast cancer screening is not effective below age 50.—JNCI 1986; 77:317-320.

A randomized trial is the best design for studying the effects of screening on mortality (*1*). Therefore, the HIP study with its 14-year follow-up (*2*) is a unique source for estimation of the effects of breast cancer screening. A number of recent randomized trials use modern mammography, e.g., in Sweden, the United Kingdom, and Canada (*3-6*), but it will take some more years before follow-up is long enough for assessing mortality effects.[6] Meanwhile, conclusions and recommendations concerning the age range for breast cancer screening will have to be based on the HIP data. The most important conclusion from previous analyses (*2, 7, 8*) is that screening results in considerable reduction in breast cancer mortality for women over 50 years of age, but that there is no clear evidence for an effect below age 50. However, an informal inspection of the 14-year follow-up results did not corroborate the conclusion of a 50-year threshold. It was therefore decided to analyze the HIP mortality results carefully, especially with respect to the problem of the age-specific effects of screening.

MATERIALS AND METHODS

In the HIP study, about 60,000 women aged 40-64 were randomly assigned to a study group or a control group. For a detailed description of the study design and of the demographic, socioeconomic, and other characteristics of the women, *see* (*9-11*).

Women in the study group were offered a screening for breast cancer by mammography and physical examination. The 67% of the women attending this first screening were offered three more screenings at annual intervals. All newly diagnosed breast cancer patients were registered, including those detected at screening. This registration was continued during some years after the fourth screening.

The analysis in this article is concerned with breast cancer patients who were diagnosed within 7 years after the start of the study. These 7 years include a period of about 3½ years after the last screening. Such a period is necessary for the control group to "catch up" with the study group, because counterparts of the screen-detected cancers will only surface in the control group after a time period known as the "lead time."

The catch-up period should not be too short. This would result in comparison of mortality among an unequal number of cases, with slower growing cancers missing in the control group. On the other hand, it should not be too long either. Otherwise, the effect of screening will become too much diluted by breast cancer cases who could not have been detected by screening, because they only came into a screen-detectable stage after the last screening round.

A catch-up period of 3½ years is considered to be appropriate for the HIP study, because after 3 years the number of cases in the control group reaches the level of the study group. Moreover, model analysis of the HIP study results indicates that well over 80% of screen-detected cases have a lead time of less than 3½ years (*12*). In case of a shorter catch-up period, e.g., of 1½ years, only about half of the cases detected at the last screening round in the study group would have surfaced by the end of the period.

The number of women in each age class is about equal for study group and control group. Therefore, number of deaths in both groups can be compared directly, without correction for a difference in the number of women in both groups. The data in the present article are calculated from the HIP statistical tape, updated until the end of 1980, but most of the relevant data can also be found in (*13*).

ABBREVIATIONS USED: CFR=case fatality rate; HIP=Health Insurance Plan of Greater New York.

---

[1] Received May 8, 1985; revised March 17, 1986; accepted April 14, 1986.

[2] Supported by grant EUR-EP-79-5 from the Dutch Cancer Foundation (Queen Wilhelmina Fund).

[3] Department of Public Health and Social Medicine, Medical Faculty, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.

[4] Netherlands Institute of Preventive Health Care, P.O. Box 124, 2300 AC Leiden, The Netherlands.

[5] We are indebted to the HIP research group (Mrs. W. Venet) for making the statistical tape of the study available. We also thank the referees for their useful comments.

[6] The first results of two Swedish projects indicate a clear reduction in breast cancer mortality (*4*).

TABLE 1.—*Basic data from the HIP study*[a]

Age class (1)	No. of women (2)	Breast cancer cases		Breast cancer deaths	
		Study group (3)	Control group (4)	Study group (5)	Control group (6)
40-44	6,500	78	83	27	39
45-49	7,240	99	93	37	43
50-54	6,935	94	103	42	54
55-59	5,920	96	95	35	43
60-64	3,650	58	69	24	33
Total	30,245	425	443	165	212

[a] Source: HIP statistical tape. (1) = 5-yr age classes; age at entry in the study; (2) = number of women (the numbers are about the same in study and control groups); (3, 4) = number of cases of breast cancer diagnosed within 7 yr after entry; (5, 6) = number of deaths with underlying cause breast cancer, among the cases given in (3, 4), within 14 yr after entry into the study.

Table 1 presents the basic data concerning number of women, number of breast cancer cases detected, and number of deaths from breast cancer. Note that the numbers in the study group include nonattenders and clinically diagnosed breast cancer cases after one or more negative screenings.

RESULTS

The effectiveness of screening in the HIP study has been measured in three ways (only deaths occurring among breast cancer cases diagnosed within 7 years after entry in the study are taken into account): *a*) the percentage reduction in number of breast cancer deaths by screening, *b*) the number of breast cancer deaths prevented per 1,000 women invited for the first screening, and *c*) the expected number of life-years gained per 1,000 women invited for the first screening.

Contrary to the other two measures, the first measure, percentage reduction in number of breast cancer deaths, is associated with the length of the catch-up period. This is a straightforward consequence of the dilutive effect of breast cancer cases that only become screen detectable after the last screening round. Thus a longer catch-up period will give rise to a smaller percentage reduction, and a shorter period will give rise to a higher percentage reductions.

Table 2 presents the results. The percentage of mortality reduction and the number of deaths prevented have been calculated from table 1. The New York life table (*14*) has been used in calculating the life-years gained. Figures differ between age groups but without a clear break at age 50. Statistical testing for homogeneity of mortality effects across age classes by chi-square goodness-of-fit is not significant. This means that the results are quite consistent with the assumption of equal effects in all classes ($P>.05$ for all three effect measures). The randomization procedure implies that the same number of breast cancer cases are to be expected in study group and control group, but chance deviations will cause differences in the observed numbers. These differences will influence the three measures of effectiveness in table 2; however, the measures can be corrected for these differences for each age class by calculating what effectiveness would have been expected in case the number of breast cancer cases in the study group would have been equal to the number actually observed in the control group.

Results are given in table 3. Comparison of tables 2 and 3 shows that the correction results in a more even distribution of reduction in mortality over the five age classes. Although the scores for effectiveness measures of screening are better than average in the youngest age class and below average in the oldest age class, statistical testing for homogeneity again reveals no significant difference for any of the three effectiveness measures both with and without corrections for differences in number of breast cancer cases.

DISCUSSION

It is beyond doubt that the results of the HIP study indicate a beneficial effect of screening on the mortality from breast cancer. The decrease in mortality in the study group, as compared with that of the control group, is significant, irrespective of the intake period (5, 7, or 10 yr after entry) or the duration of follow-up (5, 10, or 14 yr) (*2*).

The results in tables 2 and 3 indicate that there is no clear evidence that there is *no* effect of screening below age 50. The same conclusion holds when catch-up periods are chosen that are shorter or longer than the present period of 3½ years.

Tables 2 and 3 do not support the point of view that "for ages under 50 years, the HIP study does not provide evidence that screening has had an effect on mortality"

TABLE 2.—*HIP study: Effectiveness of mass screening measured by percent reduction in number of breast cancer deaths, number of breast cancer deaths prevented per 1,000 women in the study group, and number of life-years saved per 1,000 women*[a]

Age class	Mortality reduction, %	95% confidence interval	Deaths prevented per 1,000 women	95% confidence interval	Life-years saved per 1,000 women
40-44	31	−12-59	1.8	−0.6-4.4	51
45-49	14	−35-46	0.8	−1.5-4.4	20
50-54	22	−17-50	1.7	−1.0-4.6	44
55-59	19	−28-49	1.4	−1.5-4.4	29
60-64	27	−24-59	2.5	−1.6-6.8	35
Total	22	4-37	1.6	0.3-2.8	36

[a] Computations are based on the data of table 1; 95% confidence limits are given for the mortality reduction and for the prevented deaths.

TABLE 3.—*HIP study: Effectiveness of mass screening after correction (per age class) for differences in the number of breast cancer cases between study and control groups*[a]

Age class	Mortality reduction, %	95% confidence interval	Deaths prevented per 1,000 women	95% confidence interval	Life-years saved per 1,000 women
40–44	26	−17–54	1.5	−0.9–4.1	42
45–49	19	−30–51	1.2	−1.1–3.8	29
50–54	15	−24–43	1.1	−1.6–4.0	31
55–59	19	−28–49	1.4	−1.5–4.4	30
60–64	13	−38–45	1.1	−3.0–5.4	17
Total	19	1–34	1.3	0.0–2.5	31

[a] *See* table 2.

(*2*). This conclusion was first formulated in 1974 after 5-year follow-up (*15*). The use of age at diagnosis in (*15*) has rightly been questioned (*16*), but the age-specific pattern of the mortality differences was indeed irregular at that time, also when using age at entry into the study (2, table 5).

Below age 50, no difference was observed (19 deaths in study group, 20 deaths in control group). Beyond age 50, the difference (20 deaths in study group, 43 in control group) was surprisingly big in view of the fact that only 92 of the 186 cases in the study group were detected by screening and thus could have experienced any benefit at all from early detection and treatment.

When updated follow-up results became available in the following years, differences between age groups gradually disappeared (*7, 8*). However, the earlier point of view was not checked by statistically testing the homogeneity of mortality reduction across age groups. Even for the 5-year follow-up results, this test would not have resulted in a significant difference ($z = 1.34$).

An important reason for not changing the conclusion about the lack of effectiveness of screening for the women under 50 was the identification of a group of women (aged 45–49 at entry and diagnosed before the age of 50) in which there were more deaths in the study group than in the control group. Analysis of data by age at diagnosis was judged to be necessary, because if screening would not be beneficial below age 50, then women in the age class 45–49 at entry would nevertheless show benefits when they pass their 50th birthday before the last screening. Yet age at entry is the appropriate measure in the presentation of age-specific results of a randomized trial of mass screening like the HIP study. Use of age at diagnosis will inevitably result in biased comparisons, since early diagnosis causes a shift toward younger age classes for the screen-detected breast cancer cases in the study group, as compared to their control group counterparts.

This bias can explain part of the excess mortality in the group of women aged 45–49 at entry and diagnosed before age 50. It is likely that this reversal of the mortality difference between study and control group is largely due to chance effects; for example, the 14-year CFR in the control group (.40) is lower than in all other age groups (CFR = .48).

Moreover, when taking age at diagnosis into account, there appears to be one other group that differs from the average pattern. This is the group of women aged 40–44 at entry and diagnosed before age 45. Despite the above-mentioned bias that applies to this group as well, the difference in CFR is highest for this group: study CFR = .41, control CFR = .73). Note that the favorable exception also occurs in a group of women who have been diagnosed before age 50.

Neither of these findings is statistically significant because of the rather small numbers involved. Consequently, there seems to be no evidence in the HIP data for a lack of benefit below age 50, even when age at diagnosis is considered. This conclusion does not change essentially when only breast cancer cases diagnosed in the first 5 years after entry are taken into account.

The conclusion of the existence of a 50-year threshold has been criticized before (*16–18*). An important point, also made in these papers, is that the issue cannot be resolved with the results of the HIP study alone. The size of the study population chosen has been large enough for statistical confirmation of a reasonable overall decrease in mortality. In this respect, the study has been successful. However, even small differences are not supported by the data. Table 3 shows that the HIP results corroborate a hypothesis of no age dependency more than one of a considerable age gradient. The randomized controlled trials of breast cancer screening initiated in recent years in Canada, the United Kingdom, and Sweden will probably provide more conclusive evidence on the age-specific effectiveness of breast cancer screening for a number of reasons. The joint study size is much larger, some trials cover a wide age range, and intervals between screens vary between trials. Thus age dependency of sensitivity and of growth rate can profitably be assessed from these new trials.

REFERENCES

(*1*) American Cancer Society. The cancer-related health check-up. CA 1980; 30:1–39.

(*2*) SHAPIRO S, VENET W, STRAX P, et al. Ten- to fourteen-year effect of screening on breast cancer mortality. JNCI 1982; 69:349–355.

(*3*) ANDERSON I, FAGERBERG G, LUNGREN B, et al. Breast cancer screening in Sweden. The single modality approach. Radiologe 1980; 20:608–611.

(*4*) TABAR L, GAD A, FAGERBERG CJ, et al. Reduction in mortality from breast cancer after mass screening with mammography. Lancet 1985; 1:829–832.

(*5*) MILLER AB, HOWE GR, WALL C. The national study of breast cancer screening: Protocol for a Canadian randomized con-

trolled trial of screening for breast cancer in women. Clin Invest Med 1980; 4:227-258.

(6) GAD A, THOMAS BA, MOSKOWITZ M. Screening for breast cancer in Europe: Achievements, problems and future. Recent Results Cancer Res 1984; 90:179-184.

(7) SHAPIRO S. Statistical evidence for mass screening for breast cancer and some remaining issues. Cancer Detect Prevent 1976; 1:347-363.

(8) ———. Evidence on screening for breast cancer from a randomized trial. Cancer 1977; 39:2772-2782.

(9) FINK R, SHAPIRO S, LEWISON J. The reluctant participant in a breast cancer screening program. Public Health Rep 1968; 83:479-490.

(10) SHAPIRO S, STRAX P, VENET L. Periodic breast cancer screening. In: Sharp CL, Keen H, eds. Presymptomatic detection and early diagnosis. London: Pitman, 1968:203-235.

(11) FINK R, SHAPIRO S, ROESNER R. Impact of efforts to increase participation in repetitive screenings for early breast cancer detection. Am J Public Health 1972; 62:328-336.

(12) WALTER SD, DAY N. Estimation of the duration of a pre-clinical disease state using screening data. Am J Epidemiol 1983; 118:865-886.

(13) Health Insurance Plan of Greater New York. HIP Study. Evaluation of periodic breast cancer screening with mammography and clinical examination. Progress report. 1981:1-20.

(14) National Center for Health Statistics. New York State life tables: 1969-1971. Rockville, MD: NCHS, 1975.

(15) SHAPIRO S, STRAX P, VENET L, et al. Changes in 5-year breast cancer mortality in a breast cancer screening program. In: Proceedings of the seventh national cancer conference of the American Cancer Society. New York, 1974:663-678.

(16) BRESLOW L, THOMAS LB, UPTON AC. Final reports of the National Cancer Institute ad hoc working groups on mammography in screening for breast cancer. J Natl Cancer Inst 1979; 59:467-541.

(17) PROROK PC, HANKEY BF, BUNDY BN. Concepts and problems in the evaluation of screening programs. J Chronic Dis 1981; 34:159-171.

(18) MILLER AB. Screening for breast cancer. Breast Cancer Res Treat 1983; 3:143-156.