screening women over age 50. New trials must be conducted to demonstrate to what extent mammography itself contributes significantly to the benefits for this age group, as well as to demonstrate whether any improvement can be obtained for the younger age groups. If there is reason to believe that the advent of improved mammographic techniques has changed the situation from that prevailing during the HIP Study (especially for women under age 50), then the Seidman analysis could be considered to contribute to the ethical justification for conducting the appropriate randomized trials. We think planning for such trials should begin without delay, especially because a long time will be required to obtain answers to these questions. In all of our endeavors, we must always be careful that our enthusiasm for new screening, diagnostic and therapeutic techniques does not decrease our vigilance in remaining alert to the risk of iatrogenic disease.

Sylvan B. Green, M.D.
B.J. Stone, Ph.D.
Judy Chen, M.S.
Field Studies and Statistics Program,
National Cancer Institute,
Bethesda, Maryland

References

1. Seidman, H.: Screening for breast cancer in younger women: Life expectancy gains and losses. CA 27: 66-87, 1977.
2. Bailar, J.C., III.: Mammography: A contrary view. Ann. Intern. Med. 84: 77-84, 1976.
3. Upton, A.C. et al.: Report of NCI Ad Hoc Working Group on the Risks Associated with Mammography in Mass Screening for the Detection of Breast Cancer. DHEW Publication No. (NIH) 77-400. National Institutes of Health, Bethesda, Maryland, 1977.
4. Gold, R.H.: Mammography: Historical perspectives, dose, and the state of the art. Unpublished communication to the NCI Working Group to Review the NCI/ACS Breast Cancer Detection Demonstration Projects, 1977.
5. Byar, D.P. et al.: Randomized clinical trials: Perspectives on some recent ideas. N. Engl. J. Med. 295: 74-80, 1976.

Author's Reply:

The main thrust of the statement by Green et al. is that useful data on the benefits of a screening program can be obtained only through a randomized clinical trial. While I agree that clinical trials can provide the most reliable data on the long-term outcome of a program, they cannot yield all of the data necessary for an evaluation. Clinical trials are expensive in money and time; they can furnish only partial answers to many important questions, and then only if the questions can ethically be asked and a sufficient number of participants can be found to cooperate in the study. Randomized clinical trials do not escape the self-selection problems of who does not participate. Furthermore, unmeasured biases crop up as the participants self-select themselves into those who follow the specified regime and those who do not. There are also inconsistencies over a period of time as personnel change and new technologies of detection and therapy evolve. In addition, a clinical trial does not clarify whether or not length bias in the cancers detected by screening is an important consideration. In short, even clinical trials are subject to human fallibility.

Green et al. apparently agree with Breslow and his associates,1 who gave no credence to any non-clinical trial data on possible changes in benefits from screening for breast cancer subsequent to the HIP trial. I do not agree. Since we epidemiologists are far from being the sole repository of all wisdom, it was vital for us to take notice when clinicians told us over and over again that they were seeing much more very early breast cancer than was found in the HIP study. Rather than dismissing their assertions out of hand (because verifying them and assessing their significance by traditional methods was not possible), I felt there was an urgent need to make at least some provisional determinations from the available data.
These data were not merely descriptive, as downgraded by Green et al. Other authorities, also at NCI, have appreciated the valuable potential of the BCDDP data, imperfect as they are, and have invested considerable resources to collect and organize them.

The BCDDP data clearly documented what the clinicians had been saying. Many more non-invasive and small invasive cancers had been found, at ages both under and over 50, than would have been anticipated from the HIP study. I can conceive of no way in which self-selection or biases could account for this phenomenon. It is real. I believed it was important to try to approximate the magnitude of the saving of lives it might imply and therefore made an analysis fully cognizant of the limitations. It should be obvious that this was not a final answer. It was, instead, a consequence of my conviction that a tentative evaluation was a great deal better than none at all.

Green et al. regard as the central question, “What are the benefits and risks of including mammography as part of a routine screening program for detecting breast cancer in asymptomatic women?” I see this question as but one facet of the much larger problem which is, “How much of the public health dollar should be spent on screening for breast cancer, on which women, using which techniques, and how often?”

In my report1 I addressed the very basic question, “What are the benefits and risks of a total screening program offered to and accepted by a self-selected group of women?” If not for mammography, this program would never have been launched, and in this sense mammography can be credited with whatever benefits resulted.

To answer the basic question, it obviously extended the data. It is indeed ironic that Green et al. then want the data further extended to estimate the separate contributions of mammography and physical examination.

(Data do clearly show that it was mammography which was responsible for the gain in effectiveness in screening detection of the minimal cancer since the HIP trial. Also, mammography was relatively more important in detecting breast cancer among asymptomatic women than among symptomatic women.)

Green et al. evidently have an incomplete understanding of the details of my calculations and their import. For example, consider their comment, “As an additional point, it is not even clear how satisfactory are ‘incidence’ rates derived from the ACS Study, as these are based on a combination of survival data and patient questionnaires.” How trivial this comment is with respect to survival in the general population of ACS Study women may be seen from the fact that 99 percent of the 35-39 year cohort and 97 percent of the 55-59 year cohort survived the full five-year period of observation. Similarly, survival, as it relates to reporting of breast cancer cases, was trivial for these incidence rate calculations. Death reports were useful to ascertain information on women who had breast cancer diagnosed and who died during the two-year interval between follow-up questionnaires. Questionnaire reports were the basis for 90 percent of the total breast cancer cases while death reports were the sole source for but 10 percent.

In addition, I developed breast cancer incidence rates as carefully as possible in order to make estimates of mean lead time for the cancers detected by screening, and ultimately, to make estimates of what their mean life expectancy would have been subsequent to usual case findings as invasive cancers. However, for the age groups I studied using the procedures I employed, the total of such mean life expectancy plus mean lead time shows relatively little variation with mean lead time. For instance, a mean lead time of 1.50 years
is estimated in Table 4 of my paper for the cancers detected among the 45-49 year cohort at initial screening. Added to a life expectancy of 15.48 years for invasive cancers, usual case finding, interpolated from Table 6, this gives a total of 16.98 years. Suppose the mean lead time had been estimated as double, or 3.00 years. The corresponding total would have then been 17.53 years, a difference of .55 years or 3 percent. Consequently, for this calculation, the incidence rates could have varied substantially without much effect.

A misinterpretation is evidenced by Green et al. in their assertion that I assumed for the in situ cancers and for the cancers with size less than one cm. with no positive axillary nodes "just as for the more advanced categories, that without screening the 'usual invasive cancer' survival experience would apply, after a postulated mean lead time (crudely estimated using the ACS 'incidence' data)." I made no such assumption. What I did was consider the mean "usual invasive cancer" survival time plus mean lead time in relation to the total of cancers for all categories. It was well beyond the scope and intent of my paper to attempt to estimate the distributions of lead times and their variations among the different categories of cancer. Because of skewness, I would expect a weighted average of the lead times for more detailed classifications to result in a somewhat higher value of mean lead time for the total of cancers than the unweighted one I calculated. Again, as explained in the preceding paragraph, the difference would not have had much effect on my ultimate computations.

Green et al. assert that the inclusion of "current presence of lump in the breast" in my "list of so-called major risk factors" serves to cloud the issue. To them the only possible viewpoint is that "screening" is reserved exclusively for women with no symptoms or signs of breast cancer and that it becomes "diagnosis" once there are symptoms or signs. However, women with symptoms or signs comprised an important part of the women enrolled in BCDDP (and of those who participated in the HIP clinical trial) and I included such women in my analysis from the viewpoint that the purpose of screening is to examine by relatively quick procedures persons not previously known to have breast cancer in order to distinguish between those who are unlikely to have cancer and those so likely to have cancer as to warrant further intensive examination. Of the BCDDP women, ages 35-59, who reported current presence of lump in the breast at initial examination, about 10 per 1,000 women were found to have breast cancer. Whatever the semantics, this is a pick-up rate compatible with screening, not diagnosis.

There is ample evidence attesting to the generally high level of medical and technical proficiency which prevailed in the BCDDP, but there is always a significant gap between what a large-scale program (including a clinical trial program) might accomplish and what it actually does accomplish. The variability in biopsy rates at the different BCDDP centers does not "confuse the issue" but is, rather, a fruitful area to explore. It is, however, irrelevant to my report, which was concerned with the pragmatic results of the actual program. Thus, the paragraph on page 78 of my report which begins, "The pool of breast cancers detectable (and confirmable) at initial screening by the technology and medical procedures employed..." was expressly written to encompass considerations such as the variability in biopsy rates, not only from center to center, but also for younger as compared with older women.

Green et al. disapprove of the assumptions adopted for my analysis but do endorse mathematical modeling, which of necessity involves far more
extensive and elaborate assumptions. The models themselves then require results for testing. However, it was for the very reason that results were not available that I made some of the assumptions I did, and used the Surgical Adjuvant Breast Study (SABS) and the End Results Group (ERG) data for making estimates of future survival. It is not the absence of a control group which led me to use these data since a control group could hardly provide survival data at this time. Though no one can pretend to know what the actual survival will be for the BCDDP breast cancer cases, right now the ERG and SABS data are established indicators of survival for the more usual invasive cancers and there are only emerging data on the prospects for the minimal cancers.

Breast cancer is a capricious disease. Untold numbers of women have succumbed to it despite the initial belief that the cancer had been treated at a favorable stage. It is true enough that the "progression rate of in situ lesions to invasive cancer is uncertain, and the diagnosis of 'cancer' itself is not always beyond question." On the other hand, pending new therapy which can be efficacious regardless of the stage of the cancer or new methods which can invariably distinguish potentially lethal lesions, it is prudent to presume that if early detection of breast cancer leads to gains in life expectancy, then still earlier detection leads to greater gains.

Non-invasive cancers accounted for an important portion of the gains calculated in my paper, but not all of them. For example, assuming that one-half of the non-invasive cancers would not progress to invasion, about 70 percent of the gains through screening would still remain.

I believe I succeeded in obtaining an approximate answer to the question I addressed, and though the findings I reported were tentative, they were more firmly rooted than Green et al. would have one believe. With respect to the central question posed by Green et al., it is essential to look at the data as a whole. One should not consider only that part which is consistent with one's own thesis and ignore that part which is not.

There is presumed to be some risk of inducing breast cancer by the radiation used in mammography. However, this radiation has undergone an extraordinary reduction in a short period of time and this trend is continuing. Although Green et al. quote data on the Roentgens of surface radiation exposure, more important is the radiation absorbed dose within the breast. Miller et al. have recently investigated the absorbed dose for breast tissue substitute materials in relation to depth and x-ray beam quality. They found that a surface exposure of one Roentgen from a 50kVp x-ray beam (a beam typical of harder beam qualities now employed in xeromammography), resulted in a mid-breast dose of about 0.25 rad, a not unexpected finding. A surface exposure of about one Roentgen is required for xeromammograms taken with a beam of this quality; thus, an examination of a breast consisting of two xeromammographic views is accomplished with a total of about 0.5 rad to the mid-breast. However, for the molybdenum target x-ray beam, a surface exposure of one Roentgen resulted in mid-breast dose of only 0.05 rad. With this beam, a surface exposure of one Roentgen or less is required for a film-screen mammogram and two views of a breast may be taken with a dosage of about 0.1 rad to the mid-breast!

On the negative side, screening programs for breast cancer require money, personnel, facilities and equipment. Many women have to be screened to identify the few who require further investigation. Such women experience emotional and physical trauma while it
is being determined who has breast cancer and who does not.

Among the benefits of a screening program for breast cancer is the reassurance afforded women found not to have the disease. Essentially, however, the gain is concentrated among a few breast cancer patients, principally as a reduction of mortality and perhaps through the need for a less extensive surgical procedure. In general, we might expect the screening of a low breast cancer frequency group such as asymptomatic women to yield even fewer breast cancer patients who benefit. Also, the gains in the few breast cancer patients must be considerable indeed to justify the entire effort.

Table 10 of my paper indicates that the gains in life expectancy at initial screening for the group of BCDDP women termed “None plus minors only” (a group which may be regarded as asymptomatic women exclusive of those with a family history of breast cancer) were quite modest, employing a screening package which included mammography and based on an analysis which accorded benefits to minimal cancer cases which I considered plausible. These were the results measured against a period of time when mammography in general clinical practice was making but little contribution to earlier breast cancer case findings. Judged against today’s undoubtedly higher “baseline,” the gains would be less.

If one reduces the gains still further by eliminating mammography and denying a primary importance to minimal cancers, one might very well ask, “Why bother at all?”

References

1. Breslow, L.; Thomas, L.B., and Upton, A.C.: Final Reports of National Cancer Institute Ad Hoc Working Groups on Mammography Screening for Breast Cancer and A Summary Report of Their Joint Findings and Recommendations. Publication N 77.1400, U.S. Dept. of Health, Education, and Welfare, Public Health Service, National Institutes of Health, 1977.
2. Seidman, H.: Screening for breast cancer in younger women: Life expectancy gains and losses. CA 27:66-87, 1977.
3. Miller, D.W., et al.: Radiation absorbed dose in mammography. Paper presented at the 19th Annual Meeting of the American Association of Physicists in Medicine, August 1977.