To the Editor:

Readers of this Journal are surely aware of the current debate concerning the place of mammography in breast cancer screening. This subject has been discussed in a number of forums. For example, four ad hoc working groups were commissioned by the National Cancer Institute (NCI) to study this question; these are a radiation-carcinogenesis group; an epidemiology-biostatistics group and a pathology group, both of which specifically studied data developed by the Health Insurance Plan (HIP) of Greater New York; and a fourth group reviewing the ongoing Breast Cancer Detection Demonstration Projects (BCDDP). The reports of these groups are available from the Office of Cancer Communications, NCI. Others have presented mathematical models designed to investigate the effect of screening. In addition, a recent issue of this Journal contained a paper by Seidman\(^1\) which presented his analysis of life expectancy gains and losses from screening.

There are several points which we think people involved in health care should keep in mind when forming an opinion about this important matter. First, however, we would like to note that this debate, regardless of how it is eventually resolved, has already produced at least one major benefit in and of itself. This was stimulated in early 1976 when Bailar\(^2\) presented his view of the hazards of mammography, reviewed previous papers on the subject, and stressed the importance of considering the actual radiation dose delivered in practice during mammography. More recently, the report by the NCI working group chaired by Arthur Upton\(^3\) discussed in detail the risks of mammography as a function of radiation dose. Publication of such papers has led to increased attention to the monitoring and subsequent reduction of radiation dose delivered during mammography, at least for the BCDDP. Gold\(^4\) has cited data from the BCDDP comparing radiation doses delivered at the inception of each project to more recent data. With one exception, the earlier data were obtained in either May 1975 (predominantly) or April 1976; the later data, again with one exception, were obtained between December 1976 and April 1977. The surface exposure per image has decreased from a median (determined over all X-ray units then in use) of 2.7 R (range from 0.27 to 18.0) to a median of 1.2 R (range 0.2 to 2.5). The success of these efforts demonstrates the value of careful study and informed discussion of such issues relating to benefit vs. risk, and continued efforts in this direction must be encouraged.
The central question here is: "What are the benefits and risks of including mammography as part of a routine screening program for detecting breast cancer in asymptomatic women?"

The value of mammography for symptomatic women is a separate question. Clearly there may exist undisputed indications for ordering mammography for a patient with specific symptomatology (i.e., for whom cancer is already included in the differential diagnosis), indications which are not applicable to the general asymptomatic population. Seidman's inclusion, in his list of so-called major risk factors, of "current presence of lump in the breast" therefore serves to cloud the issue. Unless the lump were proven benign, its presence would be not a risk factor, but rather the disease itself. In other words, it is important not to confuse the process of screening asymptomatic populations (for the purpose of identifying individuals who presumptively have disease and who therefore require more complete diagnostic work-up) with the process of investigating the source of specific signs and symptoms (for the purpose of making a specific diagnosis and subsequently choosing the appropriate treatment plan). It is in the former setting — and not in the latter — that the role of mammography is being debated.

Furthermore, the question of interest here relates specifically to the gains associated with adding mammography to a screening program, not to the benefits of screening per se. Participants in the well-known HIP clinical trial were randomized either to a group offered screening or to a non-screened group, with mammography included as part of screening; the HIP study was not designed to test mammography per se. Similarly, Seidman's analysis of benefits pertains to the total screening program of the BCDDP, and does not determine what benefits are to be expected specifically from mammography.

As Seidman states: "BCDDP was formulated as a service demonstration project, not as a research undertaking." It was not designed as a clinical trial and cannot substitute for one. There is no control group, and without one there is no reliable way to estimate what would have happened to the BCDDP participants had they not been screened. We agree that it is useful to analyze data gathered, as part of a demonstration project. However, it is imperative that the limitations of such an analysis be kept clearly in mind, and that the results not be assumed to provide a final answer to the question at hand. If the ultimate measure of benefit is improved survival, then we should look to a well-designed and properly analyzed randomized trial to provide a convincing answer that permits quantification of how much survival difference results from inclusion of mammography as well as a measure of our confidence that the observed difference is real. Analyses such as the one presented by Seidman are based on descriptive data, and the estimates of life expectancy gained are speculative, depending heavily on the various assumptions made, both stated and implied. Unfortunately, the effects of such assumptions are not always as straightforward as one would like.

The assumptions on which the Seidman analysis depends should be carefully considered, and we note only a few of them here. First, in the absence of any control group, the various aspects of his analysis had to be based on comparisons with such diverse populations as those in the ACS Prospective Study, Third National Cancer Survey (TNCS), Surgical Adjuvant Breast Study (SABS), and the NCI End Results Group (ERG). Seidman argues that at least the first of these groups would "provide a reasonably good approximation of the population" in the BCDDP, but he then goes
on to show that the BCDDP group is "heavily overweighted with women at higher risk." If these groups differ in this obvious way, in how many other unmeasured ways do they also differ? Likewise, because of such potential biases, we cannot estimate how well or how poorly any survival predictions from SABS or ERG will relate to the BCDDP participants. As an additional point, it is not even clear how applicable "incidence" rates derived from the ACS study, as these are based on a combination of survival data and patient questionnaires. Next, Seidman's calculation of life expectancy gained by screening was performed by comparing survival probabilities based on categories of extent of disease at detection (using SABS data plus some informed guesses) with general population survival probabilities for the set of "all invasive cancers, usual case finding" (using ERG data). Seidman arbitrarily assigned the survival probabilities for the two most favorable categories (in situ; size less than one cm. with no positive axillary nodes). Yet he then assumed, for these patients 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). However, as Seidman recognizes, the progression rate of in situ lesions to invasive cancer is uncertain, and the diagnosis of "cancer" itself is not always beyond question. The more such lesions are found at biopsy incidental to the investigation of possibly suspicious findings at screening, the greater will be the apparent benefit "demonstrated" by such analysis. Note, for example, that if certain "early" lesions represent an abnormality that would never progress to invasive cancer, then treatment can have only a negative effect on survival, if it has any effect at all, but case fatality rates will look very good, because no one will die of this non-life-threatening disease. Variability in biopsy rates at different centers and in definitions of early "cancer" further confuses the issue. The net result of the various assumptions underlying the Seidman analysis is to leave us unpersuaded of the validity of the results.

Whether survival is prolonged by the inclusion of mammography in screening and by the identification and treatment of various early lesions (found more frequently in those screened with mammography) can be determined conclusively only through well-designed controlled trials.

Seidman remarks that because the clinical trial approach may involve some difficulties and requires a long period of follow-up, "alternative methods of assessment, however unsatisfying, are necessary." But keeping in mind how unsatisfactory some of these alternatives may actually be, we should not rely on them as substitutes for better ways of obtaining information. When a policy is developed that has such far-reaching public health implications, as is the case with mammographic screening of large, asymptomatic populations, we think most people would agree that firm evidence beyond assumption and conjecture is required. Certainly, analyses of currently available data are interesting and potentially informative. However, they can be misleading, especially if they encourage the belief that the answer is now known. It is not.

Two important lines of approach have been and should continue to be pursued. One is the further development of mathematical models, in which the assumptions are clearly specified and for which the implications can be directly tested. The success of such modeling will depend on the continued collection of reliable data. The second approach is the randomized controlled clinical trial. The advantages of this method of clinical investigation have been well described by Byar et al. The HIP study demonstrated a benefit from
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) 77-1400. 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 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 may 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, 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.