Chapter 16:
Modeling Cancer Natural History, Epidemiology, and Control: Reflections on the CISNET Breast Group Experience

J. Dik F. Habbema, Clyde B. Schechter, Kathleen A. Cronin, Lauren D. Clarke, Eric J. Feuer

This chapter will discuss the work of the cancer intervention and surveillance modeling network (CISNET) breast group. The discussion is structured around questions concerning the aims, results, process, and implications of the breast group experience.

WHAT WAS THE AIM?

The general aim of CISNET is modeling the impact of cancer control interventions on population trends in incidence and mortality for breast, prostate, colorectal, and lung cancer. The seven breast cancer investigations collaborated on answering the main question, “What is the impact of adjuvant therapy and screening mammography on U.S. breast cancer mortality: 1975–2000?” (1). It took 4 years of effort in model building and validation, input preparation and calibration, and producing and interpreting the results to answer the main question.

We start with discussing the formulation of the main question. Apart from adjuvant therapy and screening there have been other less conspicuous developments in treatment. These are the many small incremental improvements in preexisting treatments like surgery and anesthesia, radiation therapy, and chemotherapy. This additional contribution to survival improvement and mortality reduction was not considered in the main question. As a consequence, each participating group more or less developed its own way of dealing with this issue. Groups who fitted the incidence for 1975–2000 and predicted mortality by imposing screening and adjuvant therapy dissemination and effectiveness on the incidence ended up with a difference between observed mortality and predicted mortality. This difference can be attributed to other treatment improvements. Groups who also fitted their model to survival or mortality had to find other ways of dealing with this issue, one possibility being to attribute the mortality trend to adjuvant therapy and screening only. This issue of other treatment improvements could have been dealt with more consistently from the start and incorporated as one of the possible determinants of mortality improvement in the main question. But it would have been difficult to develop explicit inputs for these. One lesson learned in this exercise is to develop a comprehensive definition of what is and is not included in the modeling effort.

WHAT IS THE MAIN RESULT?

The seven research groups estimated the contribution of breast cancer screening and adjuvant treatment to bringing down breast cancer mortality in the year 2000, compared to the hypothetical situation in which both types of interventions had not taken place. The result has been published in the New England Journal of Medicine (2) [see also (3)]. The contribution of adjuvant treatment differs little between the models; the range is 12%–21%. The discrepancies between estimates of the contribution of screening are more pronounced: 7.5%–23%. The reactions to this latter range may vary between disappointment, i.e., stressing the large differences in view of the large body of evidence from the breast cancer screening randomized controlled trials (RCTs), to an enthusiastic conclusion that we finally have conclusive proof that breast cancer screening works on a population level because seven groups have found this result. Our interpretation is intermediate: The CISNET experience adds some evidence to the effectiveness of breast cancer screening. For each of the seven models it would have been difficult to explain all of the observed decline in breast cancer mortality between 1975 and 2000 if mammography made no impact, especially in light of the rising secular trend in mortality posited in the absence of the introduction of screening and treatment. However, we do not have seven new independent observations of its effectiveness: The use of common inputs and regular contact between groups have introduced dependencies among our results. Moreover, the results concern only one observation: the age-adjusted breast cancer mortality in the United States for 1975–2000.

Strictly speaking, the CISNET program does not comply with between-model validation principle 1 of the ISPOR Good Research Practices for modeling studies, that models should be developed independently from each other (4). On the other hand, it does comply abundantly with principle 3, that “modelers should cooperate with other modelers in comparing results and articulating the reasons for discrepancies,”, for which CISNET is applauded (4). Perhaps only comparison of final (published) results is meant, because intensive interaction between modelers on both early and later results, as in CISNET, implies dependencies in model development. Our opinion is that working together from the start onward has improved the quality of all models involved. This challenges principle 1 of the requirement of independent model development.

It is unlikely that additional RCTs will be performed to assess mammography at this time and modeling exercises such as this are increasingly being looked to as a tool to provide answers to complex questions. This work demonstrates how models can provide a piece of evidence that adds to the overall picture provided by various sources, including randomized and observational studies.
**How Should the Main Result be Communicated?**

Scientists will by and large be satisfied with the presentation of the results of seven models as given in Table 3 and Figure 3 of (2). The variability between model results is described and can be used as a starting point for discussion of determinants of these differences. Communication of the results of the seven models to policymakers is a different matter. If only one group and one model had been involved, the situation would have been straightforward. The results of the one model would be reported, with confidence intervals and a sensitivity analysis. Of course, reporting uncertainty of even one model can create uneasiness with policymakers, compared to simply reporting the best estimates of the contributions of adjuvant therapy and screening to mortality reduction. With seven models, however, we have seven series of results, as well as a complex contour of the smoothed results of each model. This graph was obtained under the assumption that the within-model variability is as large as the observed between model variability. This assumption was consistent with the within variability of the one model (M. D. Anderson) for which it was calculated (2).

That the range of results will generally become larger when using more models will paradoxically create an impression of more uncertainty and less information than when presenting the results of one model. When drawing a parallel with empirical studies, the idea of a meta-analysis of the model results is attractive. However, a meta-analysis of trial results is based on independence of the trials. The seven model results have complex and not always traceable dependencies, which makes a meta-analysis with one resulting figure difficult. The challenge for reporting multimodel results to policymakers is to keep it (nearly) as simple as reporting one-model results, but with an understanding that it is more informative and more credible. We have not yet met this challenge.

In reporting results from the base case, we chose to focus on relative results such as the percent reduction in mortality due to screening and adjuvant therapy. Relative results were felt to be less sensitive to some of the model parameters that have the greatest amount of uncertainty associated with them, most notably the background trend in incidence that would have occurred in the absence of screening. In making this decision, the group probably did not adequately consider the ease of understanding and what result would be most interesting to the public. When presenting a percent decline, the natural question is “percent of what?” Here it was a percent decline from the mortality prediction of what would have occurred in the hypothetical situation without screening and adjuvant treatment, a somewhat difficult concept to communicate. Often people would like to see the number of lives saved, an absolute result that can be more easily understood. There is a tradeoff between ease of understanding and accuracy of results that should be considered from early on in the modeling process.

**What Has Been Achieved, Apart From the Main Result?**

In the CISNET breast group, seven population models for breast cancer epidemiology and control have been developed and tested. These models can be used in addressing further epidemiological and public health questions on breast cancer. The modeling groups now go off and pursue individual interests (e.g., screening people aged 40–50 years, modeling ductal carcinoma in situ [DCIS]). Future applications will include the other envisaged uses in CISNET, to project future trends and to help determine optimal cancer control strategies. A first effort in projecting future trends has already been made, in relation to the Healthy People 2010 initiative (5). Many groups are now approaching the National Cancer Institute (NCI) who want to use the CISNET models for specific purposes—and now that the infrastructure is built, only small amounts of funding are required to do this.

The seven participating research groups have greatly increased their experience with public health modeling for cancer and have become familiar with different approaches. And quite a few junior researchers working on the project have been trained in cancer epidemiology and modeling.

The publications on the research will help to develop a quantitative orientation and understanding among clinicians and policymakers.

**How Were the Results Obtained?**

The seven groups subscribed to the same overall aims in the project proposal. It has been both a cooperative effort and a separate effort of each group.

In the semiannual meetings in which the groups presented their progress, there was much opportunity for discussion. All groups were struggling with a similar set of challenges. This made discussion and critiquing of model structure, estimation process, and simulation results deeper and more constructive than the feedback one can get from journal reviewers or audiences at conferences. These synchronous and interactive face-to-face discussions were stimulating and supported extensive cross-fertilization of ideas. The joint writing of publications added further to the detailed discussion of methods and results. CISNET was funded through NCI by a cooperative agreement mechanism, and working on a common problem definitely inspired the members to try to understand each other’s models more than if each has worked on its own modeling interests.

Importantly, during the meetings sometimes simulation results were presented that were markedly different from those of the other groups and which could be traced back to (programming) errors. Some of these errors would undoubtedly not have been identified in single-model studies and subsequently published.

The CISNET breast group has absorbed a considerable number of scientist-years of effort. Nevertheless, there has been a gain in efficiency compared with the hypothetical situation in which all groups addressed the research questions on their own. Common input was retrieved from sources once and used by all. The combined knowledge and expertise gave the group clout to get data from sources that might not have cooperated with single projects.

Some possible inefficiencies or drawbacks are the costs of the half-yearly meetings, the delays of publications because of mutual review, and embargo. In early meetings, differences in modeling terminology and philosophy led to misunderstandings and disagreements, whereas at later meetings (even though there was certainly still disagreement on many issues), the group rallied around the power of joint publication of results. In our opinion, any drawbacks are far less important than the favorable synergistic effect of the collaboration.
WHAT HAS BEEN THE ROLE OF THE COMMON INPUT IN THE RESULTS?

The models used several important common inputs and assumptions. Without these, model results would have differed much more. Using the same input instead of independently assessing and processing evidence also has the danger of a “common bias.” We do not know if such occurred. As a practical matter, lack of time and knowledge would have made it extremely difficult for each group to develop its inputs independently, and it would have made interpretation of differences between models in approaches and implementation of natural history even more complicated.

The results of an age-period-cohort (APC) analysis of breast cancer incidence for 1975–2000 were used as input by all groups (6). An APC analysis has an inherent unidentifiability problem. The parameters can be estimated only by making an arbitrary constraint on the parameter values. The chosen constraint was that there were only cohort effects but no period trends in incidence until 1982, when screening became more widespread. Under this constraint the period parameter was meant to estimate the influence of screening, whereas the cohort parameter captured other influences such as changes in risk factors over time. Although incidence had been increasing before the introduction of screening, mortality rates had been relatively constant. On the one hand, prior to 1975 (before screening and adjuvant therapy start) there were advances in surgery and general medical care, factors that would influence survival and mortality—but not incidence. In the absence of these latter advances, the background secular trend in mortality would have gone up in unison with incidence, because the cohort based risk factors would have been driving up trends in incidence for some time. Since incidence was increasing prior to 1975, while mortality was relatively flat, we have made the implicit assumption that the survival advances were just strong enough to compensate for increases in incidence. In the resulting APC estimates there is a considerable rise in incidence after 1975. With other constraints and assumptions, different incidence developments might have been obtained. This difference in turn might have had implications for the estimates of the contributions of screening and adjuvant treatment to mortality. In a sensitivity analysis, it appeared that results would not have been much different if the parameters were varied across the range of plausible identifiable models, which reassured us that the constraint that we made was not critical to the results.

As an alternative to the APC model, risk factors for breast cancer could have been modeled directly to estimate the background trend. It is not known if risk factors such as age at first child and use of hormone replacement therapy could explain the entire trend predicted by the APC model. The dissemination of mammography screening and of the two types of adjuvant therapy has been modeled based on data from National Health Interview Surveys; Surveillance, Epidemiology, and End Results (SEER) patterns of care; the Breast Cancer Surveillance Consortium; and nine SEER registries (7,8). We do not think that a different modeling of the available data on dissemination would have had a major impact on the results.

The input for death from other causes derives from the human mortality database and the National Center for Health Statistics (9) and does not leave much room for discussion. Breast cancer survival and prevalence in 1975 were based on the nine SEER registries and the Connecticut tumor registry.

For the effectiveness of adjuvant therapy, all groups have used two authoritative reviews of tamoxifen and chemotherapy (10,11). This choice limits the possibility of discrepancies between the groups. The follow-up in the review was 10 years. Survival after 10 years could be extrapolated by assuming either cure of some cancers by adjuvant treatment or delay of mortality from the cancers. Different choices of the modeling groups in this respect have somewhat influenced the results. If there had not been a meta-analysis, groups might well have used different studies with different results. Differences between modeling groups in the attribution of mortality reduction by adjuvant therapy would then have been much greater.

See (3) for more discussion on differences in mortality results among the models.

WHY THE DIFFERENCES IN RESULTS BETWEEN THE MODELS?

At an upstream level, different model results are obtained because of differences in the guiding philosophy of model building. Some models are simulations, some models are analytic, and others are hybrid. Some models are comprehensive, others parsimonious, and still others observation focused. The seven models have been compared in several publications (12–15), and we refer to these for further information and discussion. A summary table of specifications of the models is found in table 2 of (2). See the individual model chapters in this volume for more detailed descriptions.

Further downstream, the groups have used different evidence to estimate screening effectiveness. Data from many trials have been used to assess different aspects of natural history and screening parameters, which jointly have implications for the screening effectiveness. See Table 5 of (2) for a brief overview. These differences contribute to the explanation of the large differences in percentage of mortality reduction attributed to screening by the groups. But the differences are not as large as those between two reviews, one of which concluded ineffectiveness of screening, and the other considerable effectiveness (16,17).

All seven models are extremely complex, and output is determined by the joint effect of many assumptions. This is why the comparison of models on single parameters such as sensitivity or pre-clinical sojourn time and parameters related to the improvement in survival connected to earlier detection does not bring us far. It is more fruitful to compare the models on their results as applied to simple situations, so called diagnostic runs. An important diagnostic run for the breast group was the intensive screening run, in which annual screens with 100% attendance of the women started in 1976. However, even this run turned out to be too complex for an easy explanation of differences between models. Having learned from these problems, the CISNET colon group, which started later than the breast group, decided to use much simpler diagnostic runs, involving only one screen with 100% attendance. These simple runs give important and well-interpretable information on the differences in the working of the models and should be generally advised in this type of project.

Indeed, when investigating the value of intermediate outcome measures to predict mortality, the complexity of the models prohibited a good correlation for parameters like sensitivity and pre-clinical sojourn time. In fact, there were such great differences in the way these parameters were defined and used in the models.
that it was difficult even to make them more or less comparable between models (18). Also, because the use of intermediate outcome measures was not fully envisaged from the start, some groups had difficulties in producing the necessary output. Sojourn time and lead time, moreover, suffered from the differential incorporation of nonprogressing or regressing DCIS lesions, which made them incomparable between models. The intermediate outcome measures that correlated best with mortality were the incidence of advanced cancer and the program sensitivity, which is the percentage of cancers that is detected by screening (18). These two measures are the result of many interacting model parameters and have relatively close links to the final mortality result.

**What Is Special About Breast Cancer?**

Breast cancer has a privileged evidence position because many screening RCTs have been done, and large studies on the new adjuvant treatments are available. Moreover, breast cancer screening is essentially concerned with detecting invasive disease, with relatively few overdiagnosis problems, except in very old ages. But it is true that the preinvasive DCIS lesions, increasingly detectable with the newest technologies, give problems in interpretation and have been differently addressed by the modeling groups. The role of risk factors in nonfamilial breast cancer is limited. The other CISNET cancers (prostate, colorectal, and lung cancer) are more complex in one or more of these aspects.

**What Are the Implications for Model Research?**

Building seven parallel models was a little unwieldy in terms of logistics (e.g., each time we met all seven groups had to present updates) and reaching consensus. Except for lung cancer (which has five regular members and two affiliate members), each of the CISNET groups now has three regular members, including the breast cancer group. In our opinion, for important public-health questions working with several modeling groups should be considered a serious option.

For one-model research, several recommendations can be made for improving the prospect of a good quality final product:

- using a thorough model verification protocol;
- checking the consistency of results with those of simplified models;
- publishing and updating the model on a Web site according to a format such as the one used in CISNET;
- having a model-advisory group (a greater than usual commitment could perhaps be obtained when senior modelers serve as advisors on each other’s projects); and
- performing a suite of simple diagnostic runs (base cases), which helps detection of errors and eases comparison with other models.

See (4) for more suggestions for good modeling practice.

**Making Up the Balance**

Were the achievements worth the efforts? Would the scientific returns have been greater when each of the seven groups would have been working on a different problem? Throughout the chapter, pros and cons of the CISNET breast consortium have been mentioned. In summary, the main drawback is that seven groups were working during 4 years on the same research question of the impact of adjuvant therapy and screening on breast cancer mortality, although most groups also addressed several related scientific questions. The project consumed many person-years of research, including those for facilitation and coordination. The advantages of the collaboration include a much more thorough answer to the question than could have been provided by one group. The already competitive modeling groups learned a lot from each other by the intensive collaboration and interaction. Also, a consortium like CISNET is better able to access important sources of information than individual projects might be. By comparing and discussing each other’s results, errors were detected. A lesson learned is that insight in differences between models and detection of possible errors is best served by doing several really simple base cases before addressing the questions of interest. From a reader point of view, the comparative modeling approach demonstrated here gives a sense of the uncertainty associated with the choice of model and modeling assumptions. It also can give insight for interpreting other modeling results produced by these seven models. Typically, a one-model project would consider a research question and report its finding independently, leaving the reader to place the results in a larger context. And in case more groups would have addressed the same question separately, it would not have been possible to compare disparate results because of the use of completely different inputs.

Although having multiple models working together to address important issues may not always be feasible because of the time and resources involved, the breast cancer base case provided an opportunity to demonstrate the concept of comparative modeling on a scale that may not be duplicated again soon.

**Conclusion**

The CISNET breast project is a unique experience in public-health modeling. Important results for the explanation of trends in breast cancer mortality have been obtained. The models developed can be used for addressing more cancer control questions in the future. The work of the breast group has been important for the other CISNET groups. And a considerable number of CISNET-trained modelers are ready to take up new challenges. Finally, modelers in other fields can profit from the many lessons learned in this project, which we have brought together in this issue of the *JNCI Monograph*.

**References**


(12) Supplementary Appendix to (2) with a summary description of the 7 models. Available at: http://www.nejm.org.


