Statistical Methods in Medical Research 2004; **13**: 421–442

The use of modeling to understand the impact of screening on US mortality: examples from mammography and PSA testing

Eric J Feuer Statistical Research and Applications Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland, USA, **Ruth Etzioni** Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA **and Kathleen A Cronin, Angela Mariotto** Statistical Research and Applications Branch, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, Maryland, USA

Surveillance data represent a vital resource for understanding the impact of cancer control interventions on the population cancer burden. However, population cancer trends are a complex product of many factors, and estimating the contribution of any one of these factors can be challenging. Surveillance modeling is a technique for estimating the contribution of one or more interventions of interest to trends in disease incidence and mortality. In this article, we present several approaches to surveillance modeling of cancer screening interventions. We classify models as biological or epidemiological, depending on whether they model the full unobservable aspects of disease onset and progression, or models which reduce the complex process to simpler terms by summarizing portions of the disease process using mostly observed population level measures. We also describe differences between macrolevel models, microsimulation models and mechanistic models. We discuss procedures for model calibration and validation, and methods for presenting model results which are robust with respect to certain types of biased model estimates. As examples, we present several models of the impact of mammography screening on breast cancer mortality, and PSA screening on prostate cancer mortality. Both these examples are appropriate uses of surveillance modeling, even though for mammography there is extensive (although somewhat controversial) randomized trial evidence, whereas for PSA this biomarker has seen extensive use as a screening test prior to any controlled trial evidence of its efficacy. Some of the models presented here were developed as part of the National Cancer Institute's Cancer Intervention and Surveillance Modeling Network.

1 Introduction

Surveillance data – trends in cancer incidence and mortality in various subgroups of the population – constitute a vital source of information about the effects of cancer control interventions on the population disease burden. Population impact represents the final phase of cancer research as new cancer control interventions move from discovery to development to delivery. Even if an intervention has been evaluated in a randomized controlled trial (RCT), its impact on the population setting (effectiveness)

Address for correspondence: Eric J Feuer, 6116 Executive Blvd, Room 5041, MSC 8317, Bethesda, MD 20892-8317, USA. E-mail: rf41u@nih.gov

 10.1191/0962280204sm376ra

may differ from that on the trial setting (efficacy),[1] and the dissemination of the intervention to the target population may be less than complete.

Effectiveness may differ from efficacy as measured in a trial setting for several reasons. A trial represents a highly controlled environment with interventions implemented according to a strict protocol. Implementation of the intervention in the population may be quite different than that prescribed in the trial setting. Healthy individuals who tend to participate in screening trials will generally not be representative of the general population. If trials are older, the technology may have improved over time, and effectiveness may be better than efficacy. Incomplete dissemination of an intervention is a major attenuating factor in the translation of trial results to population impact. For example, suppose in a trial setting a new screening test reduces mortality by 30%, and assume its effectiveness in a population setting is 25% among those who are screened. If only 40% of the target population uses this test, then its impact would be $40\% \times 25\% = 10\%$ reduction in population mortality.

Although RCTs provide the highest level of evidence for the efficacy of new interventions, screening trials are extremely expensive and time consuming, and consequently are not always able to provide a definitive answer in a timely manner. In the case of mammography, there have been nine RCTs which have provided a rich source of information. The exact benefit of mammography, and the putative shortcomings of the various trials, is still debated, especially since these trials are not all consistent.[2,3] However, it is unlikely that additional mammography trials will be conducted. These trials were started between 1961 and 1980, and were reported between 1966 and 1999.[2] Dissemination of mammography started in the early 1980s, and accelerated in the late 1980s and early 1990s.[4] Insight into differences between trial efficacy and population effectiveness can be assessed through examination of population based observational studies.[5] Breast cancer mortality was relatively flat or slightly increasing until 1990, and has decreased 19% between 1990 and our latest 2000 data point.[6]

For PSA screening, the order of developments is reversed from that of mammography. The use of PSA in the United States rapidly disseminated starting in 1991, whereas two major trials were only first started in 1993[7] and 1994,[8] with results not expected until 2008. From 1994 to 2000, prostate cancer mortality declined by 20%.[6] These trends have led to speculation that PSA screening may be associated with survival benefit. In the absence of definitive RCT evidence, both national and international studies of population mortality trends have been cited as evidence for or against the efficacy of PSA.[9–15]

Using population disease trends to make inferences about specific interventions is a highly complex undertaking. Population trends are a product of many factors, some of which may predate the intervention of interest. These include changes in population risk profiles, improvements in diagnostic technologies, introduction of new treatments and coding changes in the way cancer deaths and diagnoses are recorded. For breast cancer, effective adjuvant multiagent chemotherapy started around 1975 with the addition of adjuvant hormonal therapy beginning around 1980,[16] and meta-analyses by the Early Breast Cancer Trialists' Collaborative Group have summarized the RCT evidence.[17,18] Strong birth cohort effects have influenced trends in breast cancer incidence,[19] although the underlying causes of the cohort trends are not completely

known. Similarly, the dissemination of PSA screening in the United States has been accompanied by changes in patterns of prostate cancer care, including an increase in the use of hormone suppression therapy for localized prostate cancer. Concurrently, incidence has been affected by declines in the use of transurethral resection of the prostate (TURP) for treating benign disease, a technology that led to the detection of many occult prostate cancers.[20] At the same time, prostate biopsy practices have changed so that the standard 4 or 6 core biopsy is now commonly replaced by a 10 or 12 core biopsy, with greater sensitivity to the presence of small tumors. Finally, PSA is used not only for screening, but also for monitoring men following a diagnosis of prostate cancer. It is likely that this use of PSA in the postdiagnostic period has led in many cases to the initiation of secondary hormone therapy years earlier than would otherwise have been the case.

How can we determine the role of our intervention of interest – here screening – in the presence of all these confounding and competing factors? Surveillance modeling is an approach designed for this purpose. A surveillance model is a set of mathematical or statistical relationships that relate an intervention (or interventions) to population level outcomes, typically incidence or mortality rates. The most basic requirement for a surveillance model is an estimate of the utilization of the intervention in the population and either an estimate or a model of its efficacy at the individual level. The surveillance model translates these input parameters into population level projections of disease incidence and mortality under the influence of the intervention. The result is a partitioning of observed disease trends into a portion attributable to the intervention and a portion attributed to 'other factors', effectively quantifying the role of the intervention in the observed trends. For some interventions, we may have a reasonable knowledge of both utilization and efficacy, whereas in other cases there may be considerable uncertainty. A key feature of a surveillance model is that it allows us to obtain estimates of the uncertainty in the model projections as a function of the uncertainty in the input parameters.

Although surveillance modeling can be applied to many types of interventions, this paper focuses on surveillance modeling as applied to the introduction of cancer screening modalities. We consider both PSA screening for prostate cancer and mammography screening for breast cancer. Despite the differences in the evidence regarding efficacy of PSA and mammography screening, surveillance modeling can provide useful answers to public health questions. For mammography, modeling can help translate the RCT evidence to the population setting, and partition the mortality decline into components attributable to treatment, screening and background risk. For PSA screening, modeling can help provide short term answers where trials are still in progress. This represents an expansion of the traditional role of surveillance data. For both mammography and PSA screening, surveillance modeling can help provide a synthesis of the current state of knowledge to allow for more informed policy decisions.

Some of the work presented herein is part of the National Cancer Institute's Cancer Intervention and Surveillance Modeling Network (CISNET) (http://www.cisnet.cancer.gov). CISNET has as its primary goal the development of models to quantify the impact of cancer control interventions at the population level. We first describe some basic principles of surveillance modeling, and then present several different models of screening mammography and PSA testing in the population. These represent a range of modeling approaches, illustrating the scope of surveillance models and the types of results that may be anticipated.

2 Principles of surveillance modeling

2.1 Types of models

A surveillance model differs from decision analysis or cost effectiveness models in that rather than representing a hypothetical cohort, it models the population, that is, a collection of birth cohorts, over a specified period of time. The model itself is a structured representation of the steps, or health states, between the intervention and the outcome.

There are several different classes of surveillance models. In *macrolevel models* (e.g., CAN*TROL[21]), standard population based statistics (birth rates, incidence, stage at diagnosis and mortality) are used to shift the population from one state to another over discrete time intervals. A set of starting conditions is specified and then a population of a certain size is run through the model. The effect of a cancer control intervention like screening is to change the transition rates between states.

In *microsimulation models*, individuals are run through the model one at a time, where at each transition a random number is generated and individual life histories are assigned accordingly. The results are then cumulated over all individuals to obtain population results. Individuals can move through discrete states, or models can be run in continuous time, and have a continuous state space, e.g., tumor size. Stochastic variation can be assessed by using a different random seed for each run.

Mechanistic models consist of a set of equations describing the relationships between key health states and/or tumor growth and metastasis. The probability distributions of output quantities of interest are then derived analytically. Parameters of these models are estimated using various statistical algorithms, e.g., maximum likelihood.

Surveillance models may also differ in the extent to which they represent the natural history of disease. This latent process underlies all models, but is typically not well understood. We characterize surveillance models as *biological* or *epidemiological*, according to their description of natural history. Biological models go beyond observable quantities to model the underlying disease onset, growth and progression. For example, a biological model may explicitly model when a tumor becomes metastatic as a function of tumor size,[22] or it may model the process of carcinogenesis.[23]

Epidemiological models *reduce* the complex process to simpler terms, usually by modeling only a portion of the entire disease process (usually the observable portion), focusing on parameters such as the incidence of diagnosed disease and the stage shift associated with screening.

Although biological models have the advantage of using parameters that characterize real biologic phenomena, they often have the disadvantage of having more parameters to estimate than epidemiological models. In biological models, many of the parameters cannot be estimated directly from observed data, and sophisticated statistical algorithms together with high quality data are required to estimate these parameters. However, these models may be subject to identifiability problems, that is, there may be several sets of parameters which fit the observed data equally well. Epidemiological models are easier to estimate, since their formulation is usually built around observed phenomena. However, epidemiological models only superficially capture the underlying process and typically lack the flexibility of biological models in expressing and estimating relations between intervention and outcome.

The following sections on mammography and PSA screening give examples of both epidemiological and biological models, and demonstrate the utility of both types of models. Simultaneous development of both types of models may be an optimal strategy, where results from one model can help inform the other in an interactive loop. Model parameterization, especially for biological models, cannot be defined in the absence of knowledge of available data sources. For example, some natural history models start with the inception of the tumor, and require data from autopsy studies;[24] whereas other models start only when a tumor is first screen detectable, and require data on the size and characteristics of screen detected cancers.[25] Another article in this issue[26] contrasts the model structure of seven different population surveillance models of breast cancers used within the CISNET consortium.

2.2 Describing model structure

Surveillance models can be complex in structure, with multiple input parameters and many underlying assumptions. Presentation of the model itself should be made as transparent as possible to facilitate interpretation of the results; however, this can be a challenge when publication space is limited. As there are no accepted standards for model reporting, descriptions of model structure should always contain the following components: 1) description of the states of the models and the possible transitions; 2) a list of the model parameters; 3) for each parameter, the source for the assumed input value and any assumption made if an estimate was not available and 4) a detailed description of how the intervention efficacy was modeled. The rationale for any sensitivity or uncertainty analysis should be well specified, especially if some parameters are highly speculative (e.g., the efficacy of an intervention when there are no trial results). If space does not permit all of the above to be included in the model report, a technical appendix should be added to provide readers with details about model structure and flow, and values used for input parameters.

2.3 Calibration versus validation

Surveillance models, particularly biological models, can be quite complex with many input parameters. One way to fit model parameters is to match model outputs with observed data. A range of values for multiple parameters can be used to identify which sets of parameter values are consistent with the observed data. This procedure is called *calibrating* the model to a specific data set. Examples of calibration data sources include randomized controlled screening and treatment trials, autopsy studies, observational studies, population based data prior to the start of the intervention (e.g., age incidence curves prior to screening) and meta-analyses. Owing to the complexity of the models and the number of parameters involved, identifiability of parameter estimates (i.e., different sets of parameters that provide the same overall fit) and overfitting of calibration data sets are persistent problems. However, prior knowledge about the possible range of the different parameters can assist in limiting these problems. After the model is calibrated, other data sets must be used to *validate* the model, that is, to evaluate whether the model produces results that match observed data not used for the calibration process.

When working with population data, data sources for calibration and validation are often limited. However, once a data set has been used for calibration, it can no longer

be used for validation. As an example, suppose our objective is to partition changes in US mortality due to various interventions (e.g., screening, treatment) and changes in risk factors. In many cases, the plethora and complexity of factors influencing observed mortality trends (e.g., the changes in surgical procedures, change in the population risk through immigration, effects of celebrity diagnoses, changes in diagnostic technologies and so on) may make validation impractical. However, just because observed mortality trends are not used for validation, does not necessarily mean they should be used for calibration. Even if the model is calibrated to observed mortality, it is still important to recognize that this does not imply that the model will produce an accurate partitioning of the overall mortality trend into its components (i.e., screening, treatment and risk factors). This is because the complex model structure may produce results that match a given set of mortality trends without actually capturing the correct underlying process. For example, the calibration process may have compensated for a missing component of the model structure by changing parameters relating to other components, or, alternatively, identifiability problems may have led to selection of a local maximum. Many different sets of parameters and their associated partitioning of the mortality trends may produce the same predictions. In this sense, calibration to complex data sources has distinct disadvantages over calibration to settings where the number of possible calibration variables is more limited. On the other hand, if observed mortality is not used for calibration, but the model validates to observed mortality, then this improves the credibility of the model produced partitioning. In addition, when a model is calibrated to observed mortality, then extension of the model to other situations (such as predicting what mortality would have been in the absence of the intervention, or extending results past the observed data) may produce less reliable results than a model that validates to observed mortality.

In spite of these disadvantages, calibration may still be an appealing option to modelers who have extensive knowledge of various input parameters (either by prior calibration steps in more controlled settings or by prior knowledge), which will limit the number and range of parameters considered in the calibration process, thus minimizing the risk of modeling the underlying process incorrectly. Calibration to observed mortality may represent a final chance, at the end of a long series of calibration steps, to 'tweak' the model to a limited extent. For example, the modeler may choose to fix all of the existing parameters, and add a new scaling parameter on the natural history submodel to reflect the impact of risk factors not explicitly modeled. The goal of calibration to observed mortality may be the overall shape of trends rather than level. Another advantage of calibration to observed mortality is that a better understanding of the attenuation of population parameters relative to randomized trial results may be achieved through this process. For example, the modeler may start with an estimate of efficacy from trials or epidemiological studies as a first pass at model inputs, and then use the observed mortality trends to calibrate the parameter so as to obtain an estimate of population effectiveness.

Intermediate solutions to using observed population trends for either calibration or validation exist. For example, it may be useful to use incidence trends for calibration, and mortality trends for validation. Alternatively, it may be useful to use mortality rates from a less recent interval for calibration, and a more recent set of years for validation.

2.4 Presenting model results

Results from surveillance models are typically presented as a comparison between incidence and mortality trends that include the intervention of interest and trends in the absence of the intervention. Some models may try to be comprehensive in modeling all of the major factors influencing disease trends, whereas others may have a more limited focus. However, because of the complexity of factors influencing trends, it is usually not possible to completely capture all of the factors which influence them. In presenting model results, it is important to do so in a way that is robust to certain types of errors and omissions in the modeling process.

Although a model may include all possible types of factors influencing trends (i.e., screening, treatment, new diagnostic technologies and changes in risk factors), it is useful to decide which are of primary interest, and which are nuisance parameters (i.e., modeled in a crude way to simply make the model complete but are not themselves of interest). We first discuss the situation where screening is the only factor of primary interest, and then extend our remarks to the situation where there are multiple factors of interest.

There are two basic approaches to represent the effect of an intervention; the *additive* method and the *relative* (or multiplicative) method. Let $M_{w/o}(y)$ be the mortality rates for year y that would have occurred in the absence of the intervention of interest (i.e., screening) but with all other factors included in the model, and $M_w(y)$ be the mortality rates with the intervention of interest and all other factors for years $y = 1, \ldots, n$.

The additive difference in trend is estimated as:

$$\text{Additive difference } (y) = \hat{M}_{w/o}(y) - \hat{M}_w(y) \tag{1}$$

It is readily apparent that if both $\hat{M}_{w/o}(y)$ and $\hat{M}_w(y)$ were both biased by an additive constant, that is, $\hat{M}_{w/o}(y) = M_{w/o}(y) + b$ and $\hat{M}_w(y) = M_o(y) + b$, then the absolute difference would still be unbiased. If, for example, prevalent cases at the start of the intervention were underestimated, then the number of deaths due to prevalent cases would also be underestimated by an additive factor both in the presence and in the absence of screening.

The relative method produces an estimate of the percent change in outcome levels with the intervention relative to levels in the absence of the intervention. The relative difference in trend is calculated as:

$$\text{Relative difference } (y) = \frac{\hat{M}_{w/o}(y) - \hat{M}_w(y)}{\hat{M}_w(y)} \tag{2}$$

If the factors not modeled or modeled incorrectly affect the outcome in a multiplicative way, that is, $\hat{M}_{w/o}(y) = bM_{w/o}(y)$ and $\hat{M}_w(y) = bM_o(y)$, then those factors cancel when calculating a relative percent change. An example of a factor that is usually hypothesized to affect mortality in a multiplicative way is the effect of birth cohort on disease risk.

Whether using additive or relative difference to estimate the effect of screening, the results can then be applied to the observed rates to produce an estimate of what

mortality rates would have been in the absence of screening. Let $M_o(y)$ be the mortality rate observed for year y. By definition, observed mortality rates are the rates which occurred in the presence of all interventions, that is, $M_w(y) = M_o(y)$. We can obtain estimates of mortality rates in the absence of screening by applying either the additive or relative estimate of the effect of screening as follows.

$$\text{Additive estimate} = (\hat{M}_{w/o}(y) - \hat{M}_w(y)) + M_o(y) \tag{3}$$

$$\text{Relative estimate} = \left[\frac{\hat{M}_{w/o}(y) - \hat{M}_w(y)}{\hat{M}_w(y)} \times M_o(y)\right] + M_o(y) \tag{4}$$

In the additive model, suppose that the estimates are biased by an additive constant. In this situation, the additive estimate (3) is an unbiased estimate of $M_{w/o}(y)$. Likewise, in the relative model, suppose that the estimates are biased by a multiplicative constant, then in this situation the relative estimate (4) is an unbiased estimate of $M_{w/o}(y)$.

One potential shortcoming of the relative estimate (2) is its awkward interpretation (i.e., percent decline due to screening relative to what would have occurred in the presence of screening). A more interpretable quantity is:

$$\frac{\hat{M}_{w/o}(y) - \hat{M}_w(y)}{\hat{M}_{w/o}(y)} \tag{5}$$

which can be interpreted as the percent decline due to screening relative to the level of mortality that would have occurred in the absence of screening. Although, like Equation (2), it is robust to multiplicative errors, the disadvantage of Equation (5) is that, unlike Equation (2), it cannot be applied to observed mortality as in Equation (4).

As by definition the intervention starts after time 0, $M_{w/o}(0) = M_w(0)$. Calibrating to $M_o(0)$, so that $M_o(0) = M_{w/o}(0) = M_w(0)$, may be appealing (and appropriate). It has the advantage that there are a limited of number of parameters to calibrate, since this set of parameters does not include those associated with the intervention. In addition, the process is robust to relative or additive discrepancies after time 0, which are generated by any difference between $M_o(0)$ and modeled mortality at time 0.

In more complicated situations, similar methods can be used to estimate the model projected effect of multiple interventions occurring over the same time period. For example, breast cancer mortality rates are influenced by several components: screening, treatment advances, birth cohort trends (usually associated with changing risk factors) and other factors. The overall goal of an analysis of mortality trends might be to partition the decline in mortality into the portion due to each component. In this case, multiple scenarios would be modeled to estimate the effect of each component separately as follows:

- $M_C(y)$ = mortality with cohort effect (no screening and treatment).
- $M_{CS}(y)$ = mortality with cohort and screening effect (no treatment).
- $M_{CT}(y)$ = mortality with cohort and treatment effect (no screening).
- $M_{CST}(y)$ = mortality with cohort, screening and treatment effect.

The effect of each intervention alone and the effect of the combined interventions could be calculated using either the additive or the relative method and used to estimate what mortality rates would have been without screening, without treatment advances or without both.

For example, estimates of the relative effects of screening and treatment are:

- Estimate of the effect of screening alone $= (\hat{M}_{CST}(y) - \hat{M}_{CT}(y))/\hat{M}_{CST}(y)$.
- Estimate of the effect of treatment advances alone $= (\hat{M}_{CST}(y) - \hat{M}_{CS}(y))/\hat{M}_{CST}(y)$.
- Estimate of the joint effect of both screening and treatment $= (\hat{M}_{CST}(y) - \hat{M}_{C}(y))/\hat{M}_{CST}(y)$.

The additive approach is similar. These relative effect estimates can be applied to observed mortality rates as shown earlier to produce estimates of what mortality would have been without screening, without treatment or without both.

An interesting result of this type of analysis is an estimate of the synergy between screening and treatment. RCTs estimate the effect of an intervention independent of other factors. In practice, the effectiveness of interventions within a population may be influenced by other factors. For example, the mortality impact of treatment advances may be modified by screen detected cases being shifted to earlier stages where survival is already favorable, and the size of any treatment advance is apt to be smaller.

In the above discussion, we have assumed that although some interventions are of primary interest, all factors influencing mortality are modeled. Suppose that certain effects are not modeled at all. For example, if cohort effects are not modeled at all, then the effect of screening alone would be $(\hat{M}_{ST}(y) - \hat{M}_{T}(y))/\hat{M}_{ST}(y)$. This estimator of this effect could still be considered a valid measure of the screening effect if we hypothesize that $\hat{M}_{ST}(y) = bM_{CST}(y)$ and $\hat{M}_{T}(y) = bM_{CT}(y)$. This may be a reasonable assumption so long as each cohort has risks which are approximately equal to that of a base cohort multiplied by a constant.

3 Applications

3.1 Understanding the impact of PSA screening on US prostate cancer mortality

PSA screening for prostate cancer represents a unique situation in which widespread use of the test has occurred ahead of definitive results about test efficacy. The test was introduced into the US population in 1988; by 1998, 38% of white men and 31% of black men over the age of 65 were being tested per year.[27] The rapid dissemination of PSA screening has dramatically increased the prostate cancer burden in the United States. Between 1987 and 1992, the age adjusted incidence of prostate cancer increased from 133.5 to 236.8 per 100 000, an increase of more than 75%.[6] Incidence has since declined, particularly distant stage incidence, as well as mortality due to the disease. Between 1991 and 2000, the age adjusted prostate cancer mortality rate decreased by 22%, from 39.3 to 30.6 per 100 000.[6]

In modeling the relationship between PSA use and population prostate cancer trends, the first task was to obtain estimates of the frequency of PSA screening beginning 1988. Unfortunately, tracking of PSA testing patterns during the early years of its use was

practically non-existent. Nationally representative population surveys such as the National Health Interview Survey (http://www.cdc.gov/nchs/nhis.htm) and the Behavioral Risk Factor Surveillance System (BRFSS) (http://www. cdc.gov/brfss/) only began including questions about PSA use in the late 1990s; prior to this time, a few states had included information on PSA utilization in their BRFSS questionnaires, but these efforts were not coordinated and sample sizes were relatively small.

In the absence of comprehensive survey data, medical claims data were utilized from the linked SEER-Medicare database (http://healthservices.cancer.gov/seermedicare/). The SEER-Medicare linked files provide information on cancer cases as well as a 5% sample of cancer free controls, residing in the SEER catchment areas. The linkage between SEER and the claims files allows for the elimination of tests conducted after a diagnosis of prostate cancer. The current version of the SEER-Medicare linkage which is available to characterize prediagnosis PSA histories covers the calendar period 1991–1999. Limitations of this data resource include lack of representative information on men below age 65, and absence of information on the reason for conducting tests. Thus, it is not possible to infer whether claims are for bonafide screening tests or for diagnostic tests conducted to confirm the presence or absence of suspected disease. The models adjust for this uncertainty by introducing nuisance parameters that effectively deflate the observed testing frequencies in an attempt to produce a more realistic approximation to true screening use.

In this section, we summarize several different modeling approaches that we have used to address questions about the role of PSA screening in US prostate cancer mortality trends. The first two models are epidemiological models – the first uses SEER-Medicare data to directly link information on the frequency of PSA testing with prostate cancer mortality, and the second links the decline in the incidence of distant stage disease with changes in disease specific mortality rates. The third model is a biological model that includes as a component a model of disease onset and progression.

3.1.1 Epidemiological model of PSA testing and prostate cancer mortality

The first model of PSA testing and prostate cancer mortality is an epidemiological model that uses PSA testing frequencies and associated cancer detection rates to generate a 'screen detected cohort'.[28] As these cases account for all changes in prostate cancer mortality, they are the focus of the model, and the relative method of presenting results is used to describe model outputs. Each member of the screen detected cohort has a date of screen diagnosis, which is the origin of their disease history in the model.

For each individual in the screen detected cohort, the model generates a date when clinical diagnosis would have occurred (t_d), a date of death in the absence of screening, and a date of death in the presence of screening. The *lead time*, which is the time from screen detection to t_d, is the first key model input, and is assumed to follow a gamma distribution with mean equal to 3, 5 or 7 years. The date of death in the absence of screening is given by $T_1 = t_d + s_1$, where s_1 is the time from clinical diagnosis to prostate cancer death in the absence of screening. Relative survival curves from SEER among cases diagnosed before the PSA era are used to approximate s_1. The date of death in the presence of screening depends on the assumed *efficacy of screening*, which is the second key parameter of the model. We denote screening benefit by a factor r,

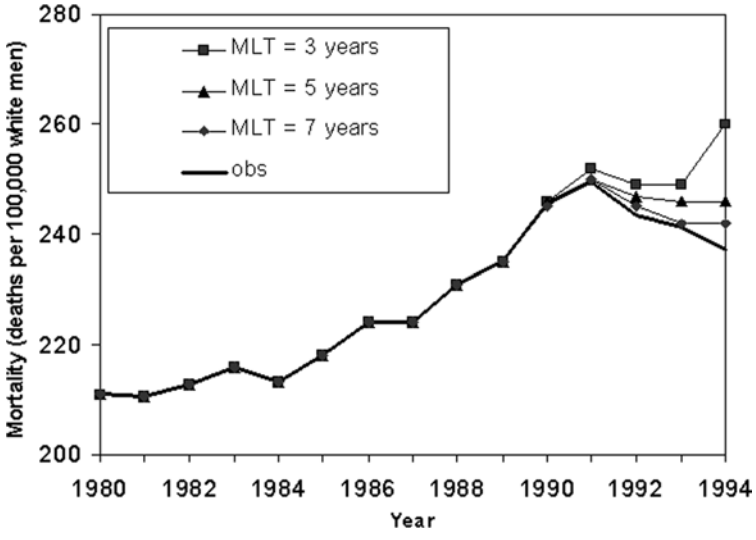

Figure 1 US prostate cancer mortality for white men aged 70–84 years; observed (1980–1994) and inflated by deaths prevented by PSA screening (1988–1994). Annual prostate cancer mortality after the date of clinical diagnosis is 50% lower among screen detected cases than among clinically diagnosed cases. Model results are given for three mean lead times (MLT) in years.

which is interpreted as the reduction in the annual risk of death from prostate cancer after the date of clinical diagnosis. Given r, the date of death in the presence of screening is given by $T_2 = t_d + s_2$, where $P(S > s_2) = P(S > s_1)^r$.

By generating for each screen detected case, a date of prostate cancer death with and without screening, the model projects the deaths prevented each year, which is simply given by the difference between the deaths in the presence and absence of PSA. The deaths prevented are converted to an age adjusted rate, and added back to the observed deaths to provide a projection of what mortality would have been in the absence of the test. On the basis of this projection, we are able to infer whether or not PSA testing plausibly accounts for the downturn in prostate cancer mortality since 1991 – if the projected mortality in the absence of PSA is increasing, for example, then this implies that PSA is responsible, for the observed decline. Figure 1 shows an example of the output of the model through 1994 – these results assume a value for r of 0.5. A value of 0.5 was chosen since it matched (after some translation) the effect size assumed in the prostate portion of the NCI's Prostate, Lung, Colorectal and Ovarian (PLCO) screening trial (see Appendix 1 of Etzioni *et al.*[28]). On the basis of the figure, we conclude that PSA screening could explain mortality declines only under fairly extreme assumptions about lead time (3 years or less) or screening efficacy ($r = 0.5$), and therefore that other factors are also likely playing a role in the observed trends.

3.1.2 *A Model of distant stage incidence and prostate cancer mortality (stage shifting model)*

The second model that we review here is a epidemiological model that quantifies the magnitude of the link between the observed decline in the incidence of distant

disease and prostate cancer mortality.[29] This model uses only information observed in the SEER database (http://seer.cancer.gov/) and a revised version of the computer model CAN*TROL.[18] The model shifts observed deficits in the number of patients with distant disease to local/regional disease.

The 'base' model assumes that a patient with screen detected local/regional disease of a given histological grade has the same prognosis as a patient with clinically detected local/regional disease of the same grade (i.e., an assumption of no length bias). Each 'shifted' distant stage case has the patient's survival from the original date of distant stage diagnosis recalculated on the basis of the pre-PSA survival distribution among local/regional cases of the same grade. A clinically plausible 'optimistic' model assumes that low grade cases shifted from distant to local/regional disease are cured. Conversely, a 'pessimistic' model assumes that those with an unfavorable grade have no benefit from being detected early. These latter two assumptions allow for the possibility that length bias is operating for cases detected early through PSA screening.

CAN*TROL2 was used to compute mortality rates and counts for each model given observed declines in the incidence of distant stage disease from 1991 to 1999. Reference mortality rates and counts were also calculated under the hypothesis that in the absence of PSA screening, the stage grade distribution during this time would have been the same as the distribution observed during the pre-PSA testing period (1980–86). Results were presented using the relative method. Given the observed declines in distant stage disease, the model projected that mortality rates would fall 18, 8 and 19% for white males and black males under the base, pessimistic and optimistic assumptions, respectively.[28]

The stage shifting model captures only mortality gains due to PSA detection that resulted in a shift from distant to local/regional disease. PSA-detected patients shifted within a given stage and cases detected by screening whose disease would not have been diagnosed in the absence of PSA (overdiagnosed cases) are not modeled. However, overdiagnosed patients do not contribute to mortality changes, and shifting of diagnosis within local/regional stage is likely to have a more delayed contribution than the shifting of diagnosis from distant to an earlier stage.

3.1.3 Biological latent disease model of PSA testing and prostate cancer mortality

The third model of PSA testing and prostate cancer mortality is a model which has at its core a model of latent natural history of prostate cancer. This begins at the onset of asymptomatic disease, and progressing through the pathological stages of prostate cancer as defined by the American Urological Association. The model is described by detail by Etzioni *et al.*[24] Transition rates from one stage to the next are based on the estimates from the Markov model of Cowen *et al.*[30] Over a specified calendar period of interest (e.g., 1980–2000), the model projects disease incidence and mortality both in the absence and presence of screening. In the absence of screening, the model selects individuals to be diagnosed with clinical prostate cancer at ages and clinical stages that match those in the SEER database prior to the advent of PSA. Each clinical case is assigned a date of prostate cancer death given by the sum of date of diagnosis and a survival time on the basis of age and stage specific relative survival curves form the pre-PSA SEER data.

In the presence of screening, the date of diagnosis is a function of the date of disease onset, PSA increases following disease onset and screening schedule. A PSA growth trajectory is created for each individual with asymptomatic onset of disease in their lifetime ($\sim$36% of the original population).[31] The rate of PSA growth is correlated with rates of transition through disease stages; specifically, the annual increase in PSA is inversely correlated with the duration of stage A1, the earliest and most lengthy pathological disease stage. If screen detection results in a shift to an earlier disease stage at diagnosis, survival is recalculated according to the new diagnosis stage. The latent disease model is still under development, but preliminary results indicate consistency with the results of the epidemiological model in that PSA screening appears to account for a portion, but not all of the observed mortality declines.

3.1.4 Summary

Table 1 summarizes the salient features of the three modeling approaches.

The models of PSA testing and prostate cancer mortality (models 1 and 3) represent a direct approach to modeling the potential impact of PSA screening on disease specific mortality trends. However, this direct approach requires information on the dissemination of PSA screening in the population as well as cancer detection rates for PSA screens, and, in the case of model 1, the lead time. In the indirect (stage shifting) approach, survival benefits under screening are modeled from the date of original clinical diagnosis. Consequently, no assumptions are necessary concerning the lead time, a quantity for which estimates vary widely.[32,33] However, the stage shifting model has its own key assumptions, namely, 1) that the relatively constant incidence of distant disease from 1973 (the beginning of SEER data collection) until 1990 would have continued in the absence of screening and 2) that the decline in distant stage disease starting in 1991 is solely due to PSA screening.

Table 1 Comparison of biological and epidemiological models quantifying the link between PSA testing and US prostate cancer mortality trends

	Epidemiological model 1	Epidemiological model 2	Biological model
Simulation point of origin	PSA-detection	Clinical diagnosis (for distant stage cases)	Disease onset
Years of simulation	1988 onwards (post-PSA era)	1990–1999	1980–2000 or as specified by user (birth cohorts begin in 1895)
Requires projection of secular trend in incidence	No	No	Yes
Requires estimate of lead time	Yes	No	No (but requires model of latent natural history)
Survival benefit owing to screening	Input relative risk	Stage-shift model	Stage-shift model

The biological model has many more parameters than the epidemiological models, but, in contrast to the epidemiological models, does not require an explicit estimate of survival benefit; this is an output of, rather than an input to, the model. One parameter required by the biological model is the secular trend in disease incidence, which is a projection of the incidence of disease that would have been expected in the absence of screening. This requires making assumptions about how other factors, like changing TURP rates, might have evolved in the absence of PSA.

Although the biological model requires far more input than the epidemiological model, it also provides far greater flexibility with respect to the user's ability to interrogate the model. For example, the biological model provides estimates of lead time and survival benefit by PSA level at diagnosis which are not possible to infer from the epidemiological model. The biological model also provides flexibility with respect to modeling additional interventions such as treatment changes or preventive approaches. These can be added to the biological model, which covers the entire population as opposed to the epidemiological model, which models only those cases detected by PSA screening. Thus, if the goal is to develop a comprehensive model of factors that may be impacting population disease trends, the biological model may be a more preferable approach. For example, none of these models currently addresses an unusual rise in mortality prior to the decline. It has been hypothesized that this rise is associated with deaths from other causes among the rising pool of prevalent cases of prostate cancer, some of which are misattributed to prostate cancer.[34] This 'misattribution bias' could be incorporated in the biological model and model 1, but would not easily fit into the more focused model 2. Ultimately, the decision about which modeling approach is most appropriate will depend on the quality of the data available to inform the model as well as the likelihood that the assumptions implicit in the model are satisfied.

3.2 Understanding the impact of mammography on US breast cancer mortality

A review of the evidence for mammography screening recently published by the Cochrane review called into questions the benefits of mammography screening for breast cancer.[2] The review examined each of the clinical trials measuring the efficacy of mammography use, identified design or analysis flaws, and classified the trial in terms of overall quality. When a meta-analysis was performed on the trials deems as acceptable quality, the review concluded that no significant reduction in breast cancer mortality was found when comparing the screened versus unscreened arms of the trials. This lead to a debate in both medical journals and the popular press on the benefits and risks associated with mammography screening for breast cancer. Another more recent review of this evidence by the US Preventative Task Force[3] led to somewhat different conclusions than the Cochrane group. With the controversy surrounding the mammography trials, there was increased interest in using population data to investigate the benefits that have been achieved in the US population since the introduction of mammography. Although analysis of population data cannot replace clinical trial evidence, it could shed some light on the many questions surrounding mammography screening. The following application presents a epidemiological modeling exercise to address this question, identifying a number of limitations related to the simple modeling of observed data and finally describes a more extensive modeling effort that is under-

way to estimate the impact of screening and treatment advanced on breast cancer mortality trends.

3.2.1 *An epidemiological modeling approach examining evidence of a stage shift*

One approach to addressing the question of the observed benefit of mammography in the US population was to use an epidemiological model that depended only on information observed in the SEER database.[6] A necessary, although not sufficient, intermediate outcome for mammography leading to a reduction in breast cancer mortality is a stage shift indicating earlier diagnosis of disease.

When a screening exam is introduced into a population, one would expect an initial increase in disease, followed by a decline and finally incidence levels would return to the trend observed before screening was introduced.[35] Between 1980 and 1999, overall breast cancer incidence rates have been increasing. However, when incidence is examined by stage and size it becomes evident that the increase in incidence is found in smaller tumors and *in situ* disease, cancers likely to have been discovered through mammography. Thus, incidence trends may be following the pattern expected after the introduction of mammography screening.

Incidence rate from nine SEER sites representing 13.9% of the US population were examined by historical stage, distant, regional, localized and *in situ* disease. Localized disease was further categorized by the size of the tumor at the time of detection (small localized defined as <2 cm and large localized defined as ≥ 2 cm) to better capture shifts in disease progression to earlier detection within the localized historical stage. Figure 2 shows age adjusted incidence trends from 1975 to 1998 by historical disease and tumor size. The increase in incidence is found to occur in smaller localized and *in situ* disease. Larger localized and regional disease shows declines in incidence, whereas distant remains fairly constant.

To further quantify how trends in stage specific disease was affected by the introduction of mammography, we used a joinpoint program that was designed for the purpose of fitting trends over time and identifying when and where changes in trend occur.[36] This software fits a linear model with random joinpoints where the linear slope changes. The statistical algorithm includes a test for the number of joinpoint, or slope changes, in a data series and estimate where those changes occur. For each of the five stage/size classifications, the joinpoint program was used to fit the log of the observed age adjusted rates. The program found the first change point for *in situ* and small localized disease to be in 1982, with an increase in slope after the 1982 joinpoint. For large localized and regional stages, the joinpoint program found an initial change in slope in 1985 and 1986, respectively, with a decrease in slope after the initial joinpoint. No joinpoint was found in distant disease. These initial joinpoints were consistent with the expected timing of a change in trend. The increase in earlier disease is quite consistent with the beginning of mammography screening dissemination in practice. As hypothesized, cases detected earlier through screening would represent a decline in later stage incidence at some later point in time. The difference between the increase in earlier stage disease and the decrease in later stage disease is related to the length of preclinical detectable disease, often referred to as lead time. Figure 2 also shows the joinpoint model fit to each data series.

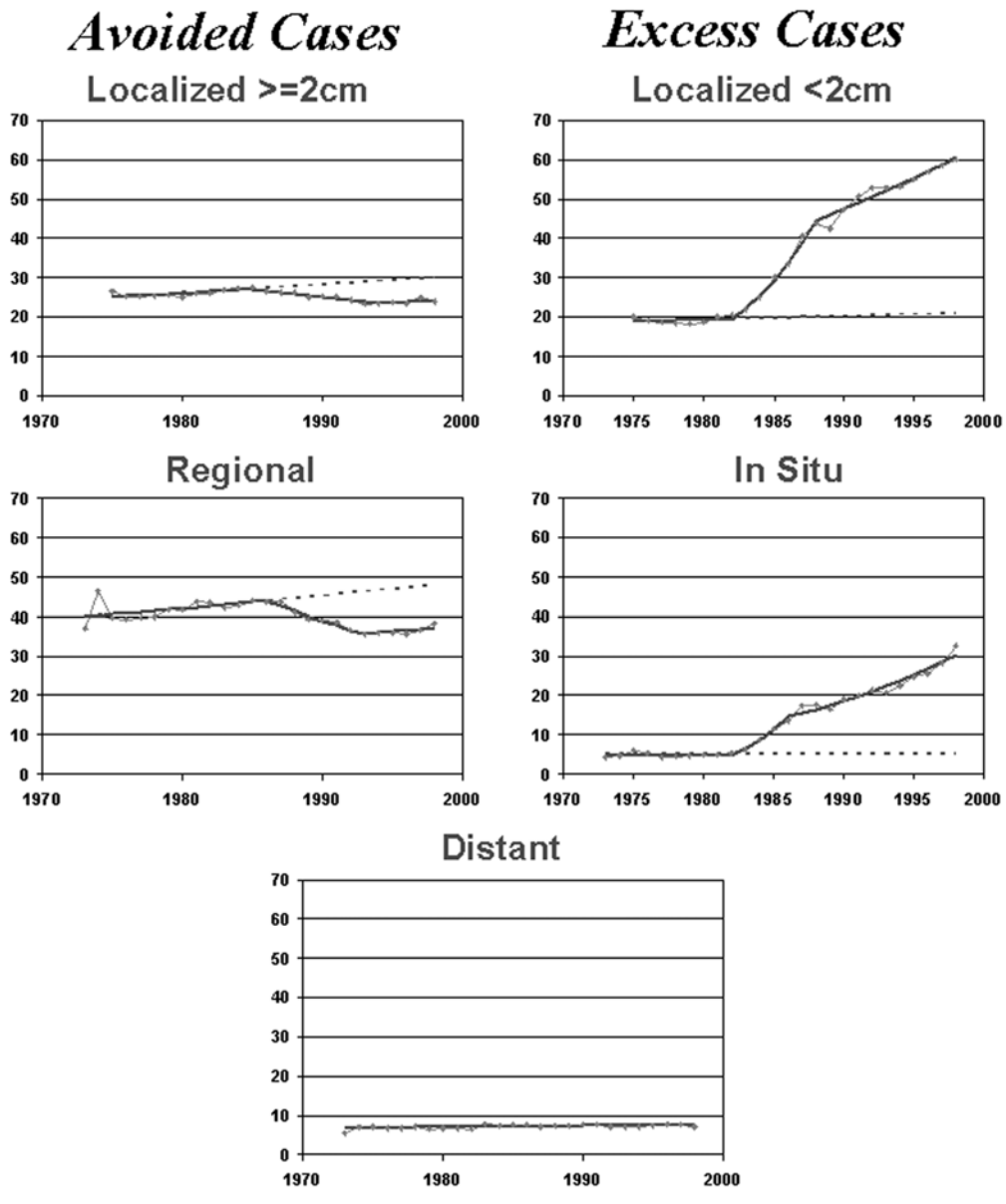

Figure 2 Avoided and excess breast cancer cases. Age adjusted rate per 100 000 women (2000 US Standard) by year. Connected dots – observed data; solid line – to join point fit; dotted line – extrapolation of initial join point trend.

The final step in quantifying the stage shift was to use the fitted joinpoint models to predict what the incidence trends would have been if screening had not been introduced. To predict incidence in the absence of screening, we extended the initial linear trend estimated in the joinpoint program. The dotted lines in Figure 2 show the

extrapolation of the initial trend from the first joinpoint to 1998. We assume that the difference between the joinpoint model and the extended initial trend represents either excess cases or cases avoided owing to screening.

One could combine the estimates of excess cases and cases avoided with stage/size specific survival to compute the net mortality benefit of the impact of the shifting size and stage distribution of breast cancer cases. The large number of excess cases with a good prognosis represents a decrement in mortality, whereas the smaller number of cases avoided with a poor prognosis represent a mortality benefit. By using stage/size specific survival estimates concurrent with when they were diagnosed, the impact of lead time on survival for early stage cases diagnosed through screening is implicitly built into the computations. If the change in the stage and size of tumors is due to introduction of mammography in the population, then this computation would provide a rough estimate of the impact of mammography on mortality trends.

In performing this simple analysis to estimate the impact of mammography, a number of limitations and areas that need further analysis became apparent. Extrapolating the initial trend may not have captured what incidence rates would have been in the absence of mammography. A strong cohort trend has been documented in breast cancer that could certainly influence results.[19] The leveling off, and even slight rise, in regional disease incidence in the mid-1990s may be due to upstaging associated with the introduction of improved technologies to find positive nodes (e.g., sentinel node biopsies[37]). In the absence of these interventions, regional disease may have continued to fall. The analysis did not try to directly model mammography use or other factors that may have influenced survival over the time period studied, such as changes in treatment. Finally, the analysis did not attempt to understand how mammography benefits survival through understanding of the natural history of disease. Although this type of epidemiological model may estimate the order of magnitude of the outcome of interest, it leaves many questions unanswered.

3.2.2 *A comparative modeling approach: the CISNET breast cancer base case*

To more comprehensively study the important question of how mammography has impacted breast cancer mortality in the US population, a group of modelers participating in the CISNET consortium considered this problem. There are seven different modelers with seven different approaches to modeling mammography in the CISNET breast cancer group. All groups agreed to participate in a controlled modeling exercise, termed the CISNET breast cancer base case. The objective of the base case is to partition the observed decline in mortality into the portion that is due to screening, the portion that is due to adjuvant chemotherapy and tamoxifen use and the portion due to other unexplained causes. An additional outcome of interest is the estimate of the interaction between screening and treatment changes, which have both occurred over the same time period.

An advantage of having a group of modelers address the same question is that it allows for a comparison of results over a wide range of modeling approaches. There is considerable uncertainty related to how to best approach modeling such a complex process. Using a number of different modeling approaches[26] allows for some insight robustness of results to various modeling assumptions. If modeling results are not

comparable, this provides an opportunity to further investigate the sources of the differences, this, in turn, will provide insight into the model parameters that have the greatest influence on results.

Figure 3 gives a schematic of how the base case will work. Although model specific inputs and assumptions vary between models, population inputs are common to all models. Data are limited on a number of population inputs, and extensive modeling was done to obtain base case inputs of interest. The dissemination of adjuvant therapy is modeled using SEER data and SEER based patterns of care studies,[16] the dissemination and patterns of repeat use of mammography are modeled using data from the National Health Interview Survey and the Breast Cancer Surveillance Consortium (Cronin KA, Yu B, Krapchow M, Miglioretti DL, Fay MP, Izmirlian G, Geller BM, Feuer EJ, personal communication) and mortality from other causes is derived using the Berkeley US cohort lifetables which are further adjusted by removing breast cancer as a cause of death (Rosenberg MA, personal communication). Because background risk (i.e., risk of disease in the absence of screening) was considered a nuisance parameter rather than of direct interest in this model, we do not explicitly model the contribution of individual risk factors to the cohort background risk trends. Instead an age period cohort model is utilized (Holford TR, Cronin K, Feuer EJ, Mariotto A, personal communication), which produces the underlying trends in risk for developing disease, sometimes referred to as cohort risk, due to the combined effect of changes in risk factors for disease over time. Controlling the population inputs will give a common starting point for all the models and facilitate comparisons between models, thus allowing the focus of the base case to be on the disease progression parameters and characteristics of the screening and treatment interventions.

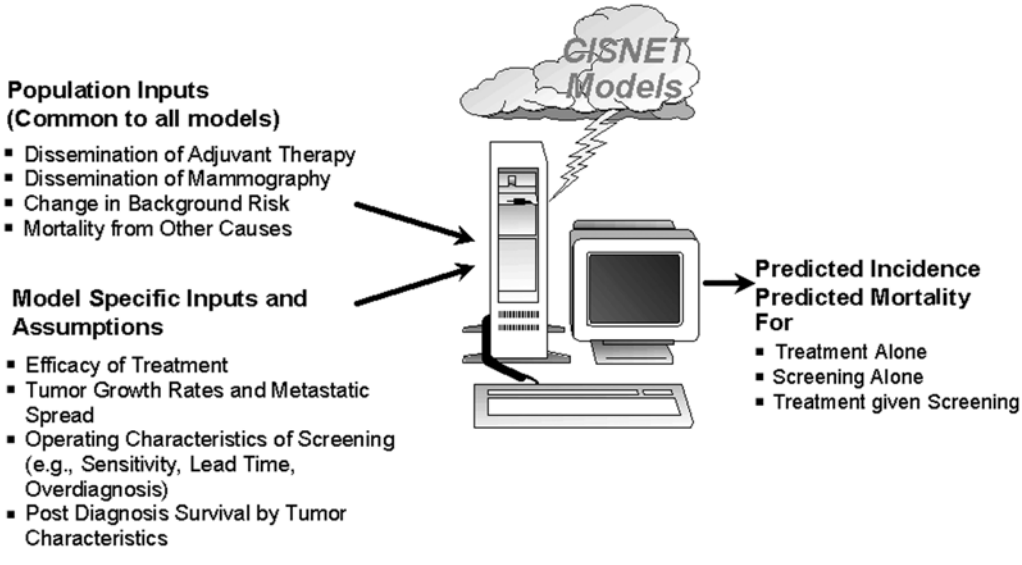

Figure 3 CISNET base question: what is the impact of mammography, adjuvant therapy, and the combination on US Breast Cancer Mortality: 1975–2000?

Models will be run several times to allow for the estimate of the contribution of screening alone, the contribution of treatment changes alone and the joint contribution of screening and treatment together to the decline in breast cancer mortality. The initial run will include only the background trend in risk for breast cancer obtained from the age period cohort model. This initial run will give a baseline for mortality trends in the absence of both screening and treatment. The difference between the baseline and mortality trends with changes in the background risk and screening will represent the effects of mammography screening in the absence of treatment. Similarly the difference between the baseline and mortality trends with changes in background risk and treatment will represent the effect of treatment alone. Finally, models will be run with changes in background risk, screening and treatment to capture the interaction between screening and treatment and their effects on breast cancer mortality. The current timeline for the breast cancer base case analysis would lead to publication of modeling results in 2005.

4 Discussion

Surveillance models are being actively used to unravel population disease trends. Modeling results should be used as part of the cumulative building blocks of science, but new results are often difficult to place in perspective with respect to earlier results. Models can be complex and difficult to report. Consequently, they are difficult to understand, and this is an obstacle to their credibility and acceptance. In this review, we have provided some metrics to classify models and present results. In addition, we have discussed the difficult area of model calibration and validation.

One way to classify models is according to whether they are epidemiological or biological. In contrast to epidemiological models, biological models go beyond observable events like screen or clinical diagnosis to model biological events such as disease onset and transition from localized to metastatic disease. In deciding between a biological or epidemiological approach, several tradeoffs must be considered. The biological approach is most flexible, but also the most demanding in terms of number of input parameters, the data resources required to estimate these parameters and the time required to develop the model. The epidemiological approach is less flexible in that the resulting models are typically designed to answer a specific question about a specific intervention, and cannot be easily extended to address multiple interventions, particularly if they interact in complex ways to influence population disease trends. As an example, when modeling primary prevention together with screening for prostate cancer, the resulting population incidence is a complex interaction between the different effects of these activities on the natural history of disease. Although prevention is expected to reduce cancer incidence, screening increases it. In addition, some preventive agents may affect the screening modality (e.g., finasteride, which dramatically reduces PSA levels).[38] Therefore, the expected changes in incidence due to prevention and screening are a result of a complex disease process, which is not captured by epidemiological models. However, epidemiological models can provide answers to critical questions within a relatively short time frame as evidenced by some of the examples presented in this article. Therefore, both types of models should be considered

as useful tools to the surveillance researcher. In some settings, simultaneous development of both types of models may be useful where results from one model can help in understanding the other in an interactive loop.

Although calibration and validation of surveillance models share many of the same issues that are relevant for other types of modeling, there are some unique issues. Because population trends in mortality are a function of many complex factors, it may be unrealistic to validate a model against them. However, just because the model will not be validated by mortality, it does not necessarily follow that it should be calibrated by observed mortality. This decision should be made taking into consideration prior calibration and validation steps, and the amount of calibration which will be allowed, to avoid overfitting to a misspecified model. In any case, methods should be used to present model results which are robust with respect to certain types of modeling omissions.

One approach to improving the credibility of surveillance modeling efforts is being utilized in the NCI sponsored CISNET consortium. The strength of having a consortium of modelers is the ability to employ a comparative modeling approach. In the past, the credibility of modeling has been marred when different models obtain widely different results that are difficult to resolve. Although each modeler has areas of individual focus, whenever possible, a common 'base' question is developed that allows for comparison across models. In these common 'base' case questions, a set of common population inputs is used across all models, and a common set of intermediate and final outputs is developed to help understand differences and similarities across models. In addition, the CISNET collaborative group has developed a state-of-the-art interactive web site, called the Model Profiler, which allows modelers to put components of their models into templates to facilitate comparisons of model structure. Each CISNET team is given a private model profile web site on which to maintain their model profile information and control what parts of their profile are shared with the rest of the consortium. As the core documentation format is the same for each group, the published profile information is readily comparable among models. The model profiling system exists to support a framework in which modelers describe their models, and interested readers read about, compare and contrast these models. After the documentation has matured, each CISNET member group can 'publish' selected documents on the main CISNET interactive web site. This site is accessible to all CISNET members and allows them to view and compare model profile documents over a range of different models. Documenting each model in a standardized way minimizes the time it takes for one group to become familiar with another's model. In the future, a user will be able to 'publish' a portion of their model profiler so that it is available to the general public and can be used as a link to be cited in publications. These steps provide a structure for comparison and a means for resolved conflicting results.

The choice of a particular modeling approach will ultimately depend on the purpose of the model, the data available for parameter estimation and the time available for model development. These models can help us be responsive to challenges due to the increasing pace of technology, provide short term answers while RCTs are still ongoing, address emerging questions while they are still being debated in the policy forum, and translate RCT evidence to the population setting. In these ways, surveillance models can

help fill the longstanding need to resolve critical policy questions which can only be answered through an indirect synthesis of available information and assumptions.

References

1 Last JM. *A dictionary of epidemiology*. Oxford University Press, fourth edition. 2001.
2 Olsen O. and Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. *The Lancet* 2001; **358**: 1340–42.
3 Humphrey LL, Helfand M, Chan BKS, Woolf SH. Breast cancer screening. Summary of the evidence. *Annals of Internal Medicine* 2002; **137**: 344–46.
4 Breen N, Wagener DK, Brown ML, Davis WW, Ballard-Barbash R. Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. *Journal of the National Cancer Institute* 2001; **93**: 1704–13.
5 Ballard-Barbash R, Taplin SH, Yankaskas BC, Ernster VL, Rosenberg RD, Carney PA, Barlow WE, Geller M, Kerlikowske K, Edwards BK, Lynch CF, Urban N, Chrvala CA, Key CR, Poplack SP, Worden JK, Kessler LG. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *American Journal of Roentology* 1997; **169**: 1001–1008.
6 Ries, LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Mariotto A, Fay MP, Feuer EJ, Edwards BK eds. *SEER cancer statistics review*, 1975–2000, Bethesda, MD: National Cancer Institute: 2003 (http://seer.cancer.gov/csr/1975–2000; 3 Last accessed 19 August 2004).
7 Gohagen JK, Prorok PC, Kramer BS, Cornett JE. Prostate cancer screening in the prostate, lung, and colorectal and ovarian cancer screening trial of the National Cancer Institute. *The Journal of Urology* 1994; **152**: 1905–909.
8 Schroder FH, Kranse R, Rietbergen J, Hoedemaeke R, Kirkels, W. The European randomized study of screening for prostate cancer (ERSPC): an update. Members of the ERSPC Section Rotterdam. *European Urology* 1999; **35**: 539–43.
9 Oliver SE, May MT, Gunnell D. International trends in prostate-cancer mortality in the "PSA ERA". *International Journal of Cancer* 2001; **92**: 893–98.
10 Shibata A, Ma J, Whittemore AS. Prostate cancer incidence and mortality in the United States and the United Kingdom. *Journal of the National Cancer Institute* 1998; **90**: 1230–31.
11 Shibata A, Whittemore AS. Re: Prostate cancer incidence and mortality in the United States and the United Kingdom. *Journal of the National Cancer Institute* 2001; **93**: 1109–10.
12 Bartsch G, Horninger W, Klocker H, Reissigl A, Oberainger W, Schonitzer D, Severi G, Robertson C, Boyle P. Prostate cancer mortality after introduction of prostate-specific antigen mass screening in the Federal State of Tyrol, Austria. *Urology* 2001; **58**: 417–24.
13 Crocetti E, Ciatto S, Zappa M. Prostate cancer: different incidence but not mortality trends within two areas of Tuscany, Italy. *Journal of the National Cancer Institute* 2001; **93**: 876–77.
14 Lu-Yao G, Albertsen PC, Stanford JL, Stukel TA, Walker-Corkery ES, Barry MJ. Natural experiment examining impact of aggressive screening and treatment on prostate cancer mortality in two fixed cohorts from Seattle area and Connecticut. *British Medical Journal* 2002; **325**: 740.
15 Coldman AJ, Phillips N, Pickles TA. Trends in prostate cancer incidence and mortality: an analysis of mortality change by screening intensity. *Canadian Medical Associate Journal* 2003; **168**: 31–35.
16 Mariotto A, Feuer EJ, Harlan LC, Wun L, Johnson K, Abrams J. Trends in use of adjuvant multi-agent chemotherapy and tamoxifen for breast cancer in the united sates: 1975–1999. *Journal of the National Cancer Institute* 2002; **94**: 1626–34.
17 Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomized trials. *The Lancet* 1998; **352**: 930–42.
18 Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomized trials. *The Lancet* 1998; **351**: 1451–67.

19 Tarone RE and Chu KC Age-period-cohort analysis of breast-, ovarian-, endometrial- and cervical-cancer mortality rates for Caucasian women in the USA. *Journal of Epidemiology and Biostatistics* 2000; **5**: 221–31.
20 Merrill RM, Feuer EJ, Warren J, Schussler N, Stephenson R. The role of transurethral resection of the prostate in population-based prostate cancer incidence rates. *American Journal of Epidemiology* 1999; **150**: 848–60.
21 Eddy DM. A computer-based model for designing cancer control strategies. *National Cancer Institute Monographs* 1986; **2**: 75–82.
22 Kimmel M, Flehinger BJ. Non-parametric estimation of the size-metastasis relationship in solid cancers. *Biometrics*, 1991; **47**: 987–1104.
23 Moolgavkar SH, Luebeck EG. Two event models for carcinogensis: biological, mathematical, and statistical considerations. *Risk Analysis*, 1990; **10**: 323–41.
24 Etzioni R, Cha R, Cowen ME. Serial prostate specific antigen screening for prostate cancer: a computer model evaluates competing strategies. *Journal Urology* 1999; **162**: 741–48.
25 van Oortmarssen GJ, Habbema JD, van der Maas PJ, de Koning HJ, Collette HJ, Verbeek AL, Geerts AT, Lubbe KT A model for breast cancer screening. *Cancer* 1990; **66**: 1601–12.
26 Boer R, Plevritis S, Clarke L. Diversity of model approaches: a review of model assumptions by the Cancer Intervention and Surveillance Modeling Network (CISNET) Breast Cancer Groups. In Uses of Stochastic Models For the Early Detection of Cancer, *Statistical Methods in Medical Research* 2004; **13**: 525–38.
27 Etzioni R, Berry KM, Legler JM, Shaw P. Prostate specific antigen testing in black and white men: an analysis of medicare claims from 1991–1998. *Urology* 2002; **59**: 251–55.
28 Etzioni R, Legler JM, Feuer EJ, Merrill RM, Cronin K, Hankey BF. Interpreting trends in prostate cancer – Part III: quantifying the link between population PSA testing and recent declines in prostate cancer mortality. *Journal of the National Cancer Institute* 1999; **91**: 1033–39.
29 Feuer EJ, Mariotto A, Merrill RM. Modeling the impact of the decline in distant stage disease on prostate cancer mortality rates. *Cancer* 2002; **95**: 870–80.
30 Cowen ME, Chartrand M, Weitzel WF. A Markov model of the natural history of prostate cancer. *Journal of Clinical Epidemiology* 1994; **47**: 3–21.
31 Etzioni R, Cha R, Feuer EJ, Davidov O. Asymptomatic incidence and duration of prostate cancer. *American Journal of Epidemiology* 1998; **148**: 775–85.
32 Pearson JD, Luderer AA, Metter EJ, Partin AW, Chan DW, Fozard JL, Carter HB. Longitudinal analysis of serial measurements of free and total PSA among men with and without prostatic cancer. *Urology* 1996; **48**: 4–9.
33 Draisma G, Boer R, Otto SJ, van der Cruijsen IW, Damhuis RA, Schroder FH, de Koning HJ. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *Journal of the National Cancer Institute* 2003; **95**: 868–78.
34 Feuer EJ, Merrill RM, Hankey BF. Interpreting trends in prostate cancer – Part II: cause of death misclassification and the recent rise and fall in prostate cancer mortality. *Journal of the National Cancer Institute* 1999; **91**: 1025–32.
35 Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine* 2000; **19**: 335–51 (correction: 2001; **20**: 655).
36 Feuer EJ, Wun L. How much of the recent rise in breast cancer incidence can be explained by increases in mammography utilization: a dynamic population approach. *American Journal of Epidemiology* 1992; **136**: 1423–36.
37 Weaver, D. Sentinal lymph nodes and breast carcinoma: which micrometastases are clinically significant? *The American Journal of Surgical Pathology* 2003; **27**: 842–45.
38 Thompson IM, Goodman PJ, Tangen CM, Lucia MS, Miller GJ, Ford LG, Lieber MM, Cespedes RD, Atkins JN, Lippman SM, Carlin SM, Ryan A, Szczepanek CM, Crowley JJ, Coltman CA Jr. The influence of finasteride on the development of prostate cancer. *The New England Journal of Medicine* 2003; **349**: 215–24.