Statistical Methods in Medical Research 1995; **4**: 18–32

Statistical models for cancer screening

CE Stevenson National Centre for Epidemiology and Population Health, Australian National University, Canberra, ACT, Australia

This paper reviews the application of statistical models to planning and evaluating cancer screening programmes. Models used to analyse screening strategies can be classified as either surface models, which consider only those events which can be directly observed such as disease incidence, prevalence or mortality, or deep models, which incorporate hypotheses about the disease process that generates the observed events. This paper focuses on the latter type. These can be further classified as analytic models, which use a model of the disease to derive direct estimates of characteristics of the screening procedure and its consequent benefits, and simulation models, which use the disease model to simulate the course of the disease in a hypothetical population with and without screening and derive measures of the benefit of screening from the simulation outcomes. The main approaches to each type of model are described and an overview given of their historical development and strengths and weaknesses. A brief review of fitting and validating such models is given and finally a discussion of the current state of, and likely future trends in, cancer screening models is presented.

1 Introduction

Screening is now regarded as a cost-effective and clinically useful approach for early diagnosis of several cancers, especially cancers of the breast and cervix. The objective of screening is the early detection of a disease where early treatment is either easier or more effective than later treatment. Cancer screening is usually carried out on asymptomatic people, since the emergence of symptoms usually indicates a later form of the disease and is likely to lead to clinical investigation.

Mathematical modelling is a useful tool in planning and evaluating screening programmes. The purpose of this paper is to describe the nature of the models used for cancer screening and how they have developed in recent years.

Extensive work has been done on quantitative theories of carcinogenesis which deal with subjects such as modelling tumour growth and response to chemotherapy. These models have little application to cancer at a population level and in particular to screening programmes and hence are beyond the scope of this study. A discussion of these models can be found in Thompson and Brown.[1] There is also a rich literature describing the use of biological models to analyse epidemiological data. These are models which describe carcinogenesis at the level of individual cells but which can be used to draw population level inferences. The most widely used of these are based on Armitage and Doll's multistage model of carcinogenesis.[2] However, these models are not appropriate for analysis of cancer screening programmes and hence are also beyond the scope of this study. Thomas[3] contains a general discussion of these models and Day[4] presents a specific discussion of the Armitage–Doll model.

This is not intended to be an exhaustive study of all modelling of cancer screening. Rather it is intended to be a description of the main approaches used in modelling and their strengths and weaknesses. Shwartz and Plough[5] present a detailed review of

Address for correspondence: CE Stevenson, Health Monitoring Division, Australian Institute of Health and Welfare, GPO Box 570, Canberra, ACT 2601, Australia.

cancer screening modelling, while Prorok discusses modelling for cervical cancer[6] and breast cancer.[7]

1.1 Why use modelling?

Randomized controlled trials (RCT) provide the most satisfactory empirical basis for evaluating screening programmes, but they are expensive and time consuming to run. Further, any one trial cannot address all the issues involved in designing a screening programme, since decisions about a programme design require joint consideration of population characteristics, resource constraints and the efficacy and effectiveness of different screening protocols.[5]

Models are one way the information on the disease and screening tests from a number of different sources, including RCTs and other clinical and epidemiological research, can be combined with known and hypothesized features of the specific population to be screened. They can be used to investigate the effect of different screening regimes on different subgroups of the population, both on disease mortality and programme costs. For example, one use of modelling has been to investigate the inclusion of different age groups in the population to be screened. They can also be used to project the future course of the disease and screening programme to evaluate the changes in costs and benefits over time.

The modelling approach does have limitations. The extra information is obtained from models only by imposing assumptions about the screening process. These include assumptions about the natural history of the disease, about the characteristics of the screening test and about the behaviour of the population under study. These assumptions can only rarely be verified, though they can be evaluated as part of the modelling process.

A further complication in making these assumptions is that the natural history of cancer is not completely understood, particularly in the asymptomatic phase which is the main focus of screening. This means that any hypothesised form of the disease model may be plausible given current knowledge but may ultimately be misleading.

1.2 Types of model

Bross *et al.*[8] proposed a classification of models used to analyse screening strategies into two types: *surface models* and *deep models*. Surface models comprise the usual statistical approach to analysis. They consider only those events that can be directly observed such as disease incidence, prevalence and mortality. Deep models, on the other hand, incorporate hypotheses about the disease process that generates the observed events. Their intent is to use the surface events as a basis for understanding the underlying disease dynamics. In cancer screening, this implies models which explicitly describe the disease natural history underlying the cancer incidence and mortality.

This modelling permits generalization from the particular set of circumstances that generated the surface events. As a result, whereas surface models provide a basis for interpreting the observable effects of screening, deep models provide an explicit basis for determining the outcomes of screening scenarios that have not been directly studied in clinical trials.[5] This study will focus on the application of deep models to cancer and population screening.

These models can be further grouped into two broad categories—those that describe the system dynamics mathematically and those which entail computer

simulation. The first of these, designated *analytic* models, uses a model of the disease to derive direct estimates of characteristics of the screening procedure and its consequent benefits. The second, designated *simulation* models, uses the disease model to simulate the course of the disease in a hypothetical population with and without screening and derives measures of the benefit of screening from the simulation outcomes.

1.3 Markov framework for modelling

Most of the cancer screening models use an illness-death model for the disease which is developed within the framework of a *Markov chain*. A sequence of random variables $\{X_k, k = 0,1,\ldots\}$ is called a Markov chain if, for every collection of integers $k_0 < k_1 < \ldots < k_n < v$,

$$\Pr\{X_v = i \mid X_{k_0}, \ldots, X_{k_n}\} = Pr\{X_v = i \mid X_{k_n}\}, \text{ for all } i.$$

In other words, given the present state (X_{kn}), the outcome in the future ($X_v = i$) is not dependent on the past ($X_{k_0}, \ldots, X_{k_{n-1}}$).

The Markov chain formulation is applied to an illness-death model in the following way.[9] The population under study is classified into n states, the first m of which are *illness states* and the remaining $n - m$ of which are *death states*. An *illness* state can be broadly defined to be the absence of illness (a *healthy* state), a single specific disease or stage of disease or any combination of diseases. In modelling cancer and screening, these states typically refer to a healthy state and preclinical and clinical stages of the disease. Here a person is considered as entering the preclinical state either at the time of carcinogenesis, or at the time when the disease is first detectable by a screening modality; and as entering the clinical state when the disease comes to clinical attention in the absence of participation in screening.

A *death* state is defined by cause of death, either single or multiple. Emigration or loss to follow up may also be treated as a death state. In modelling cancer and screening, there will typically be one death state due to death from the cancer and another due to death from any other competing cause. Entry to a terminal stage of the disease is also sometimes treated as a death state.

Transition from one state to another is determined by the *transition probabilities*, p_{ij}, where

$$p_{ij} = \Pr\{X_{k+1} = j \mid X_k = i\}, i, j = 1,2,\ldots, n;\ k = 1,2,\ldots$$

Death states are *absorbing* states, since once one reaches that state, transition to any other state is impossible (ie $p_{ij} = 0$, for $i = m + 1,\ldots,n$ and $j \neq i$). The disease model is said to be *progressive* if, once one enters the first stages of the disease, in the absence of interventions (such as screening) and competing risks, the only valid transitions are through the remaining disease stages. Because the disease is modelled using a Markov chain, the future path of an individual through the illness and death states depend only on his or her current state and the future distribution of individuals between illness and death states depends only on the present distribution and not on any past distributions.

This basic model can be varied in a number of ways. The Markov chain treats time as increasing in discrete steps corresponding to the index k. Thus a transition between

states can only occur at discrete time intervals. Most screening models extend this to allow transitions to occur in continuous time. In this case the transition probabilities for any two points in time t_1 and t_2 are

$$p_{ij}(t_1,t_2) = \Pr\{X(t_2) = j \mid X(t_1) = i\},\ i,j = 1, 2,\ldots,n$$

If $p_{ij}(t_1,t_2)$ only depends on the difference $t_2 - t_1$ but not on t_1 or t_2 separately, the model is *time homogeneous*. The simple Markov chain described above is time homogeneous. This can be varied to allow the transition probabilities to vary with time. The probabilities can also be allowed to vary with age and other relevant characteristics of the individual.

The time spent by any individual in any given state is the *sojourn time*. Many of the cancer screening models concentrate on modelling the sojourn times in preclinical states, since these are the focus of most screening programmes. Some of the model formulations allow the probability of transition out of a state to depend on the sojourn time.

2 Analytic models

A mathematical disease model with two states was first proposed by Du Pasquier,[10] but it was Fix and Neyman[11] who introduced the stochastic version and resolved many problems associated with the model.[9] Their model has two illness states and two death states. The two illness states are the state of 'leading a normal life' and the state of being under treatment for cancer. The two death states are deaths from cancer and deaths from other causes or cases lost to observation. Chiang[12] subsequently developed a general illness-death stochastic model which could accommodate any finite number of illness and death states.

Lincoln and Weiss[13] were the first to propose a model of cancer as a basis for analysing serial screening, in this case screening for cervical cancer. They did not explicitly use the Markov framework described above, but their model implicity uses a classification of the disease into two illness states—a 'healthy' state, where the disease is not detectable, and a state covering the time between when the disease is first detectable and when it is actually detected by a screening examination. They also make the simplifying assumption that symptoms never appear and all disease is detected by screening. The focus of their analysis is to derive equations to estimate the distribution of the time between the inception of the tumour and its discovery.

Zelen and Feinleib[14] proposed a simple three-state, continuous time, progressive disease model for cancer incorporating a preclinical and a clinical state. The three states are defined as:

state 1—the disease free state
state 2—the preclinical disease state
state 3—the clinical disease state.

A death state following state 3 is implicit in this model, but is not explicitly used because the analysis focuses on the preclinical state.

A preclinical state is defined as a state where a person has the disease but clinical symptoms have not been exhibited and the person is unaware of the disease. The

clinical state is taken as one having clinical symptoms of the disease. The authors assume that the preclinical disease will eventually progress to clinical disease if not detected and treated. In a modification to this basic model the authors further classify the preclinical state into two, defined as

> state 2a—a preclinical state where the disease never progresses to the clinical state (ie the sojourn time is allowed to be ∞).
> state 2b—a preclinical state where the disease is progressive and will eventually progress to the clinical state.

These are used in applying the model to cervical cancer, to allow for the possibility that some individuals with the disease in a preclinical state will never have the disease progressing to a clinical state.

Attention is confined to programmes where the individual is screened only once. The authors use the *lead time* as a measure of screening benefit. This is the time by which diagnosis is brought forward by screening. In other words, it is the time between when the individual is detected by screening in the preclinical state and when he or she would have entered the clinical state in the absence of screening. This approach has been generalized in a number of ways by subsequent authors, with most focusing on simple disease models and the estimation of specific screening characteristics.

Prorok[15,16] extended the lead time estimation to multiple screens. Blumenson[17,18,19] calculated the probability of terminal disease as a function of disease age and used this as a prognostic measure to evaluate screening strategies. Shwartz[20,21] modelled disease progression for breast cancer using tumour size and number of axillary lymph nodes involved to define states 2 and 3. He then determined screening benefit measures, from data on five year survival rate and five year disease recurrence rate for patients, as a function of tumour size and lymph note involvement.

Albert and his co-workers[22,23,24] developed a comprehensive model for the evolution of the natural history of cancer in a population subject to screening and natural demographic forces. In its general formulation the model uses Zelen and Feinleib's classification of the disease into preclinical and clinical stages, but classifies the preclinical stage into k states which are intended to correspond roughly with prognostic tumour staging schemes. It also has two death states which correspond to clinical surfacing of the disease or death from a competing risk. The model is progressive, in that once the initial disease state is entered the only valid transitions are to further disease states or to a death state, but allowance is made for staying indefinitely in any given state. This model is then applied to breast and cervical cancer. Breast cancer is modelled with two illness states, state 1 corresponding to disease with no lymphatic involvement and state 2 corresponding to disseminated disease (the contrary case). Cervical cancer is modelled with three illness states, state 1 corresponding to neoplasms *in situ*, state 2 corresponding to occult invasive lesions and state 3 corresponding to frankly invasive lesions.

The authors then impose on this model a screening strategy with a particular probability of a positive screen, depending on a person's age and disease state. Using this they derive explicit equations describing how the natural history of cancer (depicted by the distribution of numbers in each state and associated sojourn times) evolves over time in the presence of screening. These in turn are used to derive explicit equations for measures of benefit from screening in terms of the disease status, such as

the percentage reduction in the cumulative number of observed cases of late disease due to screening; and the percentage decrease in lost 'salvageables' due to screening, where a salvageable is a person who would have benefited from screening but who, in the absence of screening, progresses to a late stage of the disease before discovery.

Dublin[25,26] developed a general multi-stage disease model similar to that of Chiang[12] and applied this to breast cancer using the same two stage classification as Albert *et al.*.[23] He noted the difficulty in estimating parameter values for detailed disease models from existing data from screening programmes. His model aimed to avoid these difficulties by maintaining comparability between the model and the observable characteristics of a screened population. He did this by focusing on age and stage specific incidence and survival times in the presence and absence of screening. He derived formulae for the proportion of disease incidence which has been diagnosed earlier due to screening than it would have been in the absence of screening, and used these formulae to derive various measures of screening benefit. Dublin's model is not strictly a deep model as defined in section 1.2 above. However, although he makes no explicit hypotheses about the rate of disease progression, such hypotheses are implicit in his model.

Day and Walter[27] developed a variation on the simple three stage model which has been extensively used. The focus of this model is the sojourn time in the preclinical state, for which a probability distribution is specified. For example, Walter and Day[28] in applying the model to breast cancer, used several alternative distributions including the exponential, the Weibull and a non-parametric step function. Under the model assumptions, one may derive expressions for the anticipated incidence rates of clinical disease among groups with particular screening histories, and for the anticipated prevalence of preclinical disease found at the various screening times. One advantage of this model is that it is relatively simple to obtain approximate confidence intervals for parameter values. The model was extended by Walter and Stitt[29] to permit evaluation of survival of cancer cases detected by screening.

All of the above are progressive models but there are some forms of cancer for which the assumption of progression is not appropriate and for which some form of regression is required. These are cancers, such as large bowel cancer and particularly cervical cancer, where screening detects preinvasive or even precancerous lesions.[30]

A number of models have attempted to address this. Coppleson and Brown[31] used data on age specific clinical incidence and detection rates of a first smear to fit a four state model for cervical cancer. They found that the observed data could not be explained without allowing for regression. Albert[32] developed a variation of his earlier model for cervical cancer which allowed for regression from the carcinoma *in situ* stage back to the healthy state. Brookmeyer and Day[30] and van Oortmarssen and Habbema[33] both developed similar extensions to the Day and Walter model to divide the preclinical stage into two. The first stage allows regression to a healthy state but once a cancer reaches the second stage only progression is allowed.

The van Oortmarssen and Habbema model provides an interesting variation on the use of these models, In their study the aim of the model was not to study cancer screening directly. Rather, the model was used to study the disease dynamics and in particular to examine the epidemiological evidence for the existence of regression in pre-invasive cervical cancer.

The models described above follow a common theme of characterizing the disease as a series of states (corresponding to health, the various disease stages and death) with

people moving between the states with certain transition probabilities and/or certain sojourn times. Screening is then evaluated by superimposing on the disease process a strategy with a screen of some sensitivity. This is in contrast to the next model, due to Eddy,[34] which uses a different strategy.

Eddy's modelling strategy uses a time varying Markov framework. However, he models the interaction between the screen and the disease in his basic model. His basic model is a four stage one defined in terms of three time points. The first is a reference time point t_p. The way this is defined varies with the cancer under discussion but, as an example, for breast cancer it is the point at which the disease can first be detected by physical examination. The *occult interval* is then defined as the time interval between this and the point t_M at which the disease is first detectable by screening (e.g. by mammography). The *patient interval* is defined as the time between t_p and the time t_Π at which the patient would actually seek medical care for the lesion.

With Eddy's model t_Π, t_p and t_M can occur in any order. The important assumption is that once a disease is detectable by a screening modality (ie after t_M) then any screen using that modality will always detect the disease. This assumption replaces the normal assumption that successive screens are independent. The other two states are a 'healthy' state (which includes any preclinical disease which is still undetectable by screening) and a clinical disease state. Eddy models the probability distributions of the occult and patient intervals and uses these to derive formulae for the probabilities of discovering a malignant lesion by screening and by other methods. Eddy's model has been applied to several breast cancer screening data sets as well as to cervical cancer and gastrointestinal cancer. It has also been extended to the case where there is more than one type of screening test.[35]

Finally two recent analytic models which provide interesting variations on the basic disease model are described. The first of these is the stage shift model.[36] This model allows any number of illness stages and assumes that the effect of screening is to shift the diagnosis of a cancer from a higher to a lower stage or, within a given stage to an earlier time of diagnosis. The method of fitting this model requires a completed randomized controlled trial with equal sized intervention and control groups. It further requires that follow up has reached the point in time where comparable sets of cancer cases have accumulated in the study and control groups. This limits its applicability, but it has been successfully used to analyse breast cancer screening data.[37] The second is the peak analysis model.[38] This uses data from a randomized trial to determine the time period when the effect of screening on mortality reduction is maximum. The results of the trial can then be analysed restricting attention to that time period, providing more powerful statistical tests. For breast cancer screening, for example, this could mean excluding the mortality experience of the first few years after the initiation of screening. A disadvantage of this model is that the selection of the peak time period for the mortality comparison could be regarded as 'data-driven' and subject to the usual problems of a *post hoc* analysis.[39]

3 Simulation models

Knox[40] developed the earliest and most comprehensive simulation model. As with the analytic models, Knox uses a healthy state, a number of illness states and two death states. However, the model involves considerably more illness states, including classifying the disease as a preclinical, early clinical or late clinical cancer; and further

classifying each of these as treated or not treated, and each cancer as high or low grade. The result is a model with 26 defined states, though not every state was used for every simulation.

Knox defines a 'transition matrix' which gives the estimated transfer rates between the various pathological states, modified suitably according to the age of the individual or the duration of the state. He then simulates the evolution of the disease in a hypothetical cohort of study subjects which has similar characteristics to the population he wishes to study (which, in this case, is the adult female population of England and Wales) using the transition matrix and a standard life table to provide the risks of competing causes of death. Finally he adds details of the screening procedures to be considered, specifying the clinico-pathological states to which they apply, and their sensitivities and specificities in relation to each, and the transfers between model states which will occur following detection or non-detection. The screening policies are arranged in incremental series and the results compared with each other and with the results of providing no screening at all. This allows the appraisal of benefits and costs in both absolute and marginal terms.

This model has been applied to both cervical cancer[40] and breast cancer.[41] It illustrates one major difference between the analytic and simulation approaches. This is the greater complexity of the disease model in the simulation case. In general, simulation models are capable of considering more complex disease models and screening procedures. However, this extra complexity requires more detailed information on the disease dynamics in order to specify the model, and this information is often not readily available. Knox[42] says of his earlier work that 'The chief problem of applying the predictions stemmed from uncertainties about the clinical course of the early stages of cancer.' In this and all his subsequent analyses he simplified his model to one with only two illness states. This model is worth discussing in detail, because it provides an interesting variation on the usual modelling approach.

The major difference in this approach is in the population to which the model is applied. Whereas the usual procedure is to consider all people at risk of a cancer and to use the model to project morality with and without screening, Knox's approach is to consider only those who have died from cancer and to use the model to estimate how many would have been saved if screening had been offered. He refers to it as 'tearing down' a graph of age-distributed deaths in successive steps through the insertion of screening procedures at selected ages.[43] This means that Knox does not need to consider variations in the course of the disease such as lesions which never clinically surface or which regress to a healthy state, because all his population have, by definition, a progressive form of the disease.

The two illness states are designated A and B. During state A the disease is susceptible to early detection and full or partial cure. During state B the disease is incurable. The sojourn time in each state varies around an age specific mean. The screening procedure has a probability of detecting the lesion which rises linearly during period A, while the probability of curing the disease falls linearly during A. This model has the advantage of simplicity, which means that it is relatively easy to find plausible parameter values for it. However, this simplicity has disadvantages. The model only considers the situation of a fully established screening programme, so that it cannot be used to investigate issues surrounding setting up a new programme. Also, because it is focused on mortality reduction, it cannot be used to consider issues relating to costs of screening programmes.

Researchers at the Australian Institute of Health and Welfare have extended this approach by combining Knox's disease model with a costs model to evaluate the introduction of breast and cervical cancer screening programmes in Australia.[44,45] They have also combined the disease model with mortality projections to investigate the timing of mortality reductions due to the introduction of a breast cancer screening programme.[46]

Parkin[47] identifies a number of advantages of the usual simulation approach of transferring year by year specified proportions of a single cohort in a deterministic fashion between model states. These include the model's ability to:

- demonstrate the relationships between variables;
- explore the effects of different acceptance rates and test characteristics on outcome measures;
- examine the net cost-effectiveness of different screening policies by imputing costs to the different outcomes of screening tests; and
- explore the effect of different theoretical natural histories on the outcome of screening.

However, he also identifies some of the disadvantages of this approach. First, in practice, services have to be planned not for a single cohort over an entire life span, but for a very heterogeneous population over relatively short time periods. When a screening programme providing for testing at certain fixed ages is introduced into a community, only people younger than the starting age for the policy can possibly receive the full schedule of tests. Thus, benefits from screening will at first be small, but will increase progressively as more of the population receives a series of examinations. Further, many people will already have had previous examinations so the results of the screening policy will depend on the existing screening status of the population. This cannot be simulated by a single cohort model, nor can differences in the risk of disease in different birth cohorts.

Secondly, it may be desirable to use characteristics other than age to identify subgroups of the population for selective screening. This is less often of practical use, since such subgroups are often not readily identifiable. However, a planning model should be able to explore the effectiveness of policies involving differential screening of such sub-populations. In addition, population subgroups often have different rates of attendance at screening programmes which may be correlated with different disease risks.

Finally, in real life, screening programmes do not exist in isolation from the rest of the health care system. Much screening activity can take place outside a screening programme. Most models usually treat this activity as 'diagnostic' and ignore it. However, a planning model should take account of all relevant screening activity. Parkin proposes instead a *microsimulation* approach. Here the life histories of individual members of a population are simulated. The population in his model has the demographic make up of that of England and Wales and its size is governed by two considerations: (i) the computer time involved in microsimulation of very large populations and (ii) the need for reliable results in a stochastic simulation of relatively rare events.

Each individual is characterized by his or her values for a set of variables which will be used in simulating demographic events, disease natural history or screening programmes. The values of these variables are updated annually using sets of

conditional transition probabilities (e.g. the probability of childbirth given age, marital status and initial parity). The occurrence of a transition is decided by comparing the relevant probability against a randomly generated number. The disease model used is of the standard form. The cancer is conceptualized as having a number of states, including a health state and a death from other causes state.

There is considerable flexibility in modelling screening programmes and, since the model follows individuals, it is possible to simulate contacts with the health care system and the 'incidental' screening which occurs on such occasions.

Parkin's microsimulation model was developed specifically for cervical cancer screening, but a group working at Erasmus University in the Netherlands has developed a general modelling framework for microsimulation modelling of cancer screening called MISCAN (MIcrosimulation SCreening ANalysis).[48,49] Strictly speaking, MISCAN is not itself a model, but rather a model generator—a package which can generate and calculate a variety of these microsimulation models. The MISCAN approach, like Parkin's model, is based on the actual structure of a population as it develops in a given country at a particular time. The mass screening programme under consideration is taken as starting in a particular year and finishing in a particular year. Standard demographic techniques are used to project the study population to a year well after the nominated end of the programme. This allows for both the introduction of the programme to be modelled and the effects after the end of the programme to be followed up.

The basic structure of the cancer model is similar to Knox's earlier one with a detailed classification of clinical and preclinical cancer states, though it uses a smaller number of states. The interaction between the disease model and the screening programme is designed to allow projection and screening and treatment costs as well as cancer mortality and morbidity.

MISCAN has been widely used to analyse breast and cervical cancer screening programmes.

4 Model fitting and validation

Eddy[50] proposed four levels of validation for mathematical models. These are:

- first order validation: which requires that the structure of the model make sense to people who have a good knowledge of the problem;
- second order validation: which involves comparing estimates made by the model with the data which was used to fit the model;
- third order validation: which involves comparing the predictions of the model with data which was available when the model was fitted but was not used in the estimation of model parameters;
- fourth order validation: which involves comparing the outcomes of the model with observed data when applied to data generated and collected after the model was built (for example, data from a previously unobserved screening programme).

In this section I shall discuss model fitting and validation in the framework of these levels.

First order validation is generally not difficult to accomplish. The conceptualization of cancer as a series of preclinical and clinical stages is virtually universally accepted as a reasonable characterization of the disease. Problems may arise when the details of the

disease stages are specified but generally a wide variety of model formulations are plausible within the constraints of the limited knowledge of preclinical cancer. Second order validation highlights one of the central problems with this sort of deep model. This is the difficulty of relating available data directly to model parameters. The mismatch between the data available, either from screening trials or other sources, and the model data requirements for parameter estimation has been recognized from the beginning of this type of modelling. Lincoln and Weiss[13] note, for example, that 'Here we can do no more than introduce plausible forms for the different functions involved and plausible values for the parameters' (p 188). They go on to describe the difficulties in relating available data to the mathematical functions on which their model is based. This is a recurring problem in modelling cancer for screening and to some extent affects all the models described in this study.

Some of the analytic models have developed methods of estimating model parameters using standard statistical estimation approaches. Dublin,[26] for example, structured his model so that it could directly use the data from screening trials, though as a consequence his model relates less to disease natural history than the others . Louis *et al.*[24] derived non-parametric models for the probability distributions specified in their model and proposed the use of maximum likelihood methods to fit them. Day and Walter[27] used both parametric and non-parametric functions for their preclinical sojourn time and suggest either maximum likelihood methods or least-squares criteria to fit them. However, many of the analytic models and all of the simulation models proceed in a more *ad hoc* fashion by varying their disease natural history and model parameters until their models closely reproduce existing data.

Knox[41] gives an example of how this *ad hoc* fitting operates, in fitting his earlier model to breast cancer screening data. He describes fitting the natural history data thus: 'A statement of the natural history of the disease process must be provided in the form of a "transition matrix" which gives estimated transfer rates between the various pathological states, modified suitably according to the age of the woman or the duration of the state. This set of values is adjusted iteratively until an output is produced which matches available data on incidence, prevalence and mortality. If, as sometimes happens, more than one natural history statement is capable of mimicking these facts, then the natural history will have to be treated as one of the uncertainties. Subsequent runs will then have to be repeated for a range of natural history alternatives, and each prediction of results will be conditional upon the accuracy of the natural history used' (pp 17–18).

Parkin[47] provides an example of just such an uncertainty about natural history, with the final model including three different natural histories as alternatives.

This approach to model fitting has the disadvantage that, particularly for models with a large number of unknown parameters, the fit of the predicted values may be close to the observed data, whether or not the model is in any sense valid. However, fitting the model to a number of independent data sets simultaneously and validating it against each of these data sets, as was done for example by van Oortmarssen *et al.*,[49] provides some protection against this possibility.

Third order validation is usually made difficult by the lack of data. Generally most available data are used in determining the parameters of the model.[5] This is particularly true for breast cancer models.

The only real data source for fitting models for breast cancer screening are the screening studies and in particular the randomized controlled trials. The first major

study was the Health Insurance Plan of New York study (HIP).[51] This programme started screening in 1963. Subsequent studies were not started for another 10 years, with the Utrecht Screening Program[52] starting screening in 1974 and the Swedish Two-county Randomised Trial starting in 1977.[53] This means that many of the models only had access to the HIP data. Screening technology has changed significantly since the HIP programme began,[49] so when later studies became available, they could not be directly compared with the HIP programme and, in any case it is questionable whether models based only on HIP data are directly relevant to modern screening. Because of the long time before mortality benefits from screening are fully apparent, models fitted using solely data from later studies have only appeared relatively recently[49] and, at least in their published form, have generally not addressed the issue of third order validation. However, as more screening programmes are implemented, more data should become available for third order validation.[44]

Eddy[50] recognized that fourth order validation is possible only in rare cases. However, there are at least two examples of studies which use models in a way that could be called fourth order validation, coincidentally both using Eddy's own model. Verbeek *et al.*[54] compare predictions from Eddy's model for breast cancer to data from a mammography screening programme in Nijmegen. The authors note that the comparison does not suggest too good a fit. However, this is only a preliminary study and further validation work remains to be done. Eddy[55] compares his model for cervical cancer with a later independent analysis of empirical data. In this case the model appears to predict accurately the effect of different cervical cancer screening policies on outcomes that are important for policy decisions.

5 Current state and future directions

The problem of model validation and its effect on the credibility of model based results is still a barrier to their wider use. Nevertheless there are a number of areas where modelling can make a uniquely important contribution to our current understanding of screening.

In the absence of specific randomized controlled trials, modelling remains the only effective way of evaluating different screening regimes. For example, the inclusion of women aged between 40 and 50 in a mammography screening programme is still a contentious issue with no international consensus on the effectiveness of screening for these women.[44] While it could be argued that decisions on screening these women should not be made in the absence of reliable evidence on the presence or absence of the benefits, in practice governments are already developing screening programmes, and modelling plays an important role in guiding policy makers.

Modelling also has a crucial role to play in assessing the cost effectiveness of screening programmes. Even for cheap and easily available screening technologies, organized mass screening programmes are the best way to ensure that the benefits of screening are fully realized.[45] Modelling is not necessary only in order to plan these programmes, but no funding body is likely to fund such a programme without proper cost-effectiveness studies, and modelling is the only practical way to derive the necessary estimates of future benefits and costs.

Miller *et al.*[39] best summarize the current situation when, in discussing some recent models, they say 'It is clear that these, and other models already developed or under consideration, may enhance our understanding of the natural history of screen-

detected lesions and the process of screening. However, they require validation with the best available data, which is preferably derived from randomised trials, before they could be extrapolated in ways that might guide policy decisions. As such data become available, assumption-based models need to be modified to incorporate this extra information, in order to improve the extrapolations needed to make policy.' (p 768). While analytic models have a role in investigating specific facets of the disease and screening process (e.g. van Oortmarssen and Habbema[33]), the more comprehensive simulation models, and particularly the microsimulation models, seem best suited to the overall assessment of costs and effectiveness in screening programmes and the investigation of different screening regimes. However, the challenge in using the simulation approach is to derive disease and screening models which are sufficiently complex to model all relevant aspects of screening, but sufficiently simple to enable interpretable second order validation.

Acknowledgements

I thank Professor Susan Wilson and Drs Jane Matthews and Terry O'Neil for their helpful comments on an earlier draft of this paper

References

1 Thompson JR, Brown BW. *Cancer modeling.* New York, Basel: Marcel Dekker, 1987.

2 Armitage P, Doll R. Stochastic models for carcinogenesis. In:*Fourth Berkeley Symposium of Mathematical Statistics and Probability.* California: University of California Press, 1961: 19–27.

3 Thomas D. Use of biologic models to analyze epidemiologic data – Application: stochastic models of carcinogenesis, gene-environment interactions, and breast and lung cancer. In: Dwyer JH, Feinleib M, Lippert P, Hoffmeister H eds. *Statistical models for longitudinal studies of health.* New York, Oxford: Oxford University Press, 1992: 351–366.

4 Day NE. The Armitage–Doll multistage model of carcinogenesis. *Statistics in Medicine* 1990; **9**: 677–79.

5 Shwartz M, Plough A. Models to aid in planning cancer screening programs. In: Cornell RG ed. *Statistical methods for cancer studies.* New York, Basel: Marcel Dekker, 1984: 239–416.

6 Prorok PC. Mathematical models and natural history in cervical cancer screening. In: Hakama M, Miller AB, Day NE eds. *Screening for cancer of the uterine cervix.* IARC Scientific Publication 76, 1986: 185–98.

7 Prorok PC. Mathematical models of breast cancer screening. In: Day NE, Miller AB eds. *Screening for breast cancer.* Toronto: Hans Huber, 1988: 95–109.

8 Bross IDJ, Blumenson LE, Slack NH, Priore RL. A two disease model for breast cancer. In: Forrest APM, Bunkler PB eds. *Prognostic factors in breast cancer.* Baltimore: Williams and Wilkins, 1968: 288–300.

9 Chiang CL. *An introduction to stochastic processes and their applications.* Huntington, New York: Robert E Krieger Publishing Company, 1980.

10 Du Pasquier. Mathematische Theorie der Invaliditatsversicherung. *Milt Verein Schweiz Versich-Math* 1913; **8**: 1–153.

11 Fix E, Neyman J. A simple stochastic model of recovery, relapse, death and loss of patients. *Human Biology* 1951; **23**: 205–41.

12 Chiang CL. A stochastic model of competing risks of illness and competing risks of death. In: Gurland J ed. *Stochastic models in medicine and biology.* Madison: University of Wisconsin Press, 1964: 323–354.

13 Lincoln T, Weiss GH. A statistical evaluation of recurrent medical examination. *Operations Research* **1964; 12**: 187–205.

14 Zelen M, Feinleib M. On the theory of screening for chronic diseases. *Biometrika* 1969; **56**: 601–14.

15 Prorok PC. The theory of periodic screening I: Lead time and proportion detected. *Advances in Applied Probability* 1976; **8**: 127–43.

16 Prorok PC. The theory of periodic screening II: Doubly bounded recurrence times and mean lead time and detection probability

estimation. *Advances in Applied Probability* 1976; **8**: 460–76.
17 Blumenson LE. When is screening effective in reducing the death rate? *Mathematical Biosciences* 1976; **30**: 273–303.
18 Blumenson LE. Detection of disease with periodic screening: Transient analysis and application to mammography examination. *Mathematical Biosciences* 1977; **33**: 73–106.
19 Blumenson LE. Compromise screening strategies for chronic disease. *Mathematical Biosciences 1977;* **34**: 79–94.
20 Shwartz M. An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease. *Cancer* 1978; **41**:1550-64.
21 Shwartz M. A mathematical model used to analyse breast cancer screening strategies. *Operations Research* 1978; **26(6)**: 937–55.
22 Albert A, Gertman PM, Louis TA. Screening for the early detection of cancer–I. The temporal natural history of a progressive disease state. *Mathematical Biosciences* 1978; **40**: 1–59.
23 Albert A, Gertman PM, Louis TA, Liu S-I. Screening for the early detection of cancer–II. The impact of the screening on the natural history of the disease. *Mathematical Biosciences* 1978; **40**: 61–109.
24 Louis TA, Albert A, Heghinian S. Screening for the early detection of cancer–III. Estimation of disease natural history. *Mathematical Biosciences* 1978; **40**: 111–44.
25 Dublin N. Benefits of screening for breast cancer: Application of a probabilistic model to a breast cancer detection project. *Journal of Chronic Diseases* 1979; **32**: 145–51.
26 Dublin N. Predicting the benefit of screening for disease. *Journal of Applied Probability* 1981; **18**: 348–60.
27 Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984; **40(1)**: 1–14.
28 Walter SD, Day NE. Estimation of the duration of a preclinical state using screening data. *American Journal of Epidemiology* 1983; **118**: 865–86.
29 Walter SD, Stitt LW. Evaluating the survival of cancer cases detected by screening. *Statistics in Medicine* 1987; **6**: 885–900.
30 Brookmeyer R, Day NE. Two-stage models for the analysis of cancer screening data. *Biometrics* 1987; **43(3)**: 657–69.
31 Coppleson LW, Brown B. Observations on a model of the biology of carcinoma of the cervix: a poor fit between observations and theory. *American Journal of Obstetrics and Gynecology* 1976; **125**: 153–57.
32 Albert A. Estimated cervical cancer disease state incidence and transition rates. *Journal of the National Cancer Institute* 1981; **67**: 571–76.
33 van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. *British Journal of Cancer* 1991; **64(3)**: 559–65.
34 Eddy DM. *Screening for cancer: theory, analysis and design.* Englewood Cliffs, New Jersey: Prentice-Hall, 1980.
35 Eddy DM, Nugent FW, Eddy JF *et al.* Screening for colorectal cancer in a high-risk population. *Gastroenterology* 1987; **92**: 682–92.
36 Connor RJ, Chu KC, Smart CR. Stage-shift cancer screening model. *Journal of Clinical Epidemiology* 1989; **42**: 1083–95.
37 Chu KC, Connor RJ. Analysis of the temporal patterns of benefits in the Health Insurance Plan of Greater New York Trial by stage and age. *American Journal of Epidemiology* 1991; **133(10)**: 1039–49.
38 Baker SG, Connor RJ, Prorok PC. Recent developments in cancer screening modeling. In: Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC eds. *Cancer screening.* Cambridge: Cambridge University Press, 1991: 404–18.
39 Miller AB, Chamberlain J, Day NE, Hakema M, Prorok PC. Report on a workshop of the UICC Project on evaluation of screening for cancer. *International Journal of Cancer* 1990; **46(5)**: 761–69.
40 Knox EG. A simulation system for screening procedures. In: McLachlan G ed. *Future and present indicatives, problems and progress in medical care, ninth series.* Oxford, London: Nuffield Provincial Hospitals Trust, 1973: 17–55.
41 Knox EG. Simulation studies of breast cancer screening programmes. In: McLachlan G ed. *Probes for health.* London, New York, Toronto: Oxford University Press, 1975: 13–44.
42 Knox EG. Evaluation of a proposed breast cancer screening regimen. *British Medical Journal* 1988; **297**: 650–54.
43 Knox EG, Woodman CBJ. Effectiveness of a cancer control programme. *Cancer Surveys* 1988; **7(3)**: 379–401.
44 Australian Health Ministers' Advisory Council. *Breast cancer screening in Australia: future directions.* Canberra: Australian

Institute of Health: Prevention Program Evaluation Series No 1. AGPS, 1990.
45 Australian Health Ministers' Advisory Council. *Cervical cancer screening in Australia: options for change.* Canberra: Australian Institute of Health: Prevention Program Evaluation Series No 2. AGPS, 1991.
46 Australian Institute of Health and Welfare. *Australia's health 1992: the third biennial report of the Australian Institute of Health and Welfare.* Canberra: AGPS, 1992: 83.
47 Parkin DM. A computer simulation model for the practical planning of cervical cancer screening programmes. *British Journal of Cancer* 1985; **51(4)**: 551–68.
48 Habbema JDF, Lubbe JThN, van der Maas PJ, van Oortmarssen GJ. A computer simulation approach to the evaluation of mass screening. In: van Bemmel *et al.* ed. *MEDINFO 83. Proceedings of the 4th World Conference on Medical Informatics.* Amsterdam: North Holland, 1983: 1222.
49 van Oortmarssen GJ, Habbema JD, van der Maas PJ *et al.* A model for breast cancer screening. *Cancer* 1990; **66(7)**: 1601–12.
50 Eddy DM. Technology assessment:the role of mathematical modeling. In: Institute of Medicine ed. *Assessing medical technologies.* Washington DC: National Academy Press, 1985: 144–53.
51 Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten to fourteen year affect of screening on breast cancer mortality. *Journal of the National Cancer Institute* 1982; **69**: 349–55.
52 Collette HJA, Day NE, Rombach JJ, de Waard F. Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case control study. *Lancet* 1984; **I**: 1224–26.
53 Tabar L, Fagerberg G, Duffy S, Day NE. The Swedish two county trial of mammography screening for breast cancer: recent results and calculation of benefit. *Journal of Community Health* 1989; **43**: 107–14
54 Verbeek ALM, Straatman H, Hendriks JHCL. Sensitivity of mammography in Nijmegen women under age 50: some trials with the Eddy model. In: Day NE, Miller AB eds. *Screening for breast cancer.* Toronto: Hans Huber Publishers, 1988: 29–38.
55 Eddy DM. The frequency of cervical cancer screening. Comparison of a mathematical model with empirical data. *Cancer* 1987; **60(5)**: 1117–22.