

Work in progress: a paper to be presented at the combined CES-HESG meeting, Paris, January 2004

Modelling screening programmes: unobservable obstacles?

Jon Karnon, Liddy Goyder, Alan Brennan, Jim Chilcott, John Brazier

ScHARR, University of Sheffield

## Introduction

The purpose of screening is to classify persons as being at either greater or lesser risk of having a particular condition. There are well established criteria for screening programmes,<sup>1</sup> which have been adapted by the UK National Screening Committee (NSC) to guide the provision of screening programmes in general, as well as to inform the specification of accepted screening programmes. The criteria address four broad factors: the condition, the test, the treatment, and the screening programme.

The condition, it is stated, must be important, and the natural history and epidemiology must be understood. The screening test should be simple, safe, precise and acceptable to the general population, and there should be a defined diagnostic process following a positive test. Treatment for screen-detected disease should lead to better outcomes than treatment provided at the point of clinical diagnosis. Regarding the screening programme as a whole, it is stated that plans for monitoring the programme should be defined, adequate staffing and facilities should be available to cope with expected demand, and that the programme should provide value for money, as compared to other areas of medical expenditure.

The last criterion states the need for screening to be cost-effective, and implicitly, if screening is cost-effective, that the most cost-effective form of screening should be implemented. Each of the preceding criteria describe factors that must be defined in order to estimate the cost-effectiveness of screening, or that will enable the confirmation of cost-effectiveness, i.e. monitoring. The criteria recognise the need for cost-effectiveness to be defined in terms of a generic outcome measure to enable comparison with other areas of medical expenditure, which in practice requires the estimation of quality adjusted life years (QALYs).

The criteria also state that “there must be evidence from high quality Randomised Controlled Trials that the screening programme is effective in reducing mortality or morbidity”. If such data were available, one could assess the cost-utility of alternative screening programmes alongside

the relevant trials. However, there are areas in which no clinical trial evidence is available, for example, cervical cancer screening, and it is always the case that trial evidence is not available to inform all possible specifications of a screening programme, for example, with respect to the eligible population, the combination of screening tests and the interval between screening rounds. Let alone issues over the generalisability of trial results.

It is, therefore, necessary to use some form of modelling to inform cost-utility analyses of screening programmes. This paper reports the initial findings from a review of modelling studies of screening programmes, focussing on the area of cancer screening. The aim of the paper is to describe, and assess, alternative approaches to different aspects of model-based cost-utility analyses of cancer screening programmes.

The methods section describes the search strategies used to identify relevant papers, and the review process. The results section firstly presents a summary of the modelling studies identified in two of the main cancer screening areas: breast and colorectal cancer. Alternative model structures, and the associated assumptions regarding the disease process, screening and the interaction between the two, are then compared. The next section describes and compares three broad approaches to model parameterisation that were identified during the review of cancer screening models.

## Methods

Detailed searches of Medline, Embase, the economic evaluations databases produced by the NHS and the Office of Health Economics, and specialist operational research sources such as INFORMS and International Abstracts in Operational Research (IAOR) informed the first stages of the review. Various permutations of search terms describing the context (e.g. screening), the process (e.g. economic evaluation) and the method (e.g. modelling) were used in each database.

Abstracts from all retrieved references were reviewed by two researchers (JK, EG). Differences were discussed and resolved amicably. The set of retrieved papers included a lot of economic evaluation papers for which it was difficult to exclude the possibility modelling, though it was noted that the reference lists of such papers might be useful sources of additional modelling studies.

Additional studies have been identified from the reference lists of the retrieved papers, and from responses to the project website [www.], which asked for additions to the list of the screening studies identified from the initial literature searches.

A preliminary review of all identified cancer papers was undertaken using a data extraction form as illustrated in Figure 1.

**Figure 1 Data extraction form for preliminary review**

Disease area	
Disease incidence	How have they measured incidence? Have they measured symptomatic or asymptomatic incidence?
Disease detection	How has the stage of disease at detection (either screen- or clinical-diagnosis) been determined: sources and methods
Disease progression	How is disease progression described, both pre-clinical and post-clinical: sources and methods
Test characteristics	How are the test characteristics estimated: sources and methods
Treatment effectiveness	How is treatment effectiveness at screen- or clinical-detection estimated?
Model type	What type of model is used, does it have any special features, e.g. modelling resource constraints or dynamic incidence/prevalence?
Outcome measurement	What outcome measure is used, any comments on original estimation methods.
Cost issues	Any comments on original estimation methods.
Uptake	Do differential uptake rates affect cost-effectiveness, by what mechanism, e.g. is risk or cost a function of uptake?
Comments	Any other comments.

The data collected during the preliminary review was used to refer back to studies addressing particular aspects of the model-based cost-utility process.

## Results

The results of the review are presented in three sections: a summary of model structures and assumptions in identified breast and colorectal cancer screening modelling studies, a comparison of model structures, and a comparison of broad parameterisation approaches.

### Breast cancer model structures

Breast cancer probably has the simplest natural history pathway as there is only a single type of non-invasive cancer (ductal Carcinoma in Situ (dCIS)), which does not regress. Breast cancer appears to have been the first area of cancer screening to which modelling studies were applied, and a range of analytical models, based on a series of inter-relating equations, were published in the 1970s. Since then, a range of other modelling techniques have been applied to breast cancer screening.

Some newer studies model breast cancer as a single state,<sup>2,3</sup> from which patients may only progress to death. These models apply relative risk reductions for breast cancer mortality to observed mortality rates for non-screened populations.

Both Allen et al<sup>4</sup> and Mandelblatt et al<sup>5</sup> use decision trees to model individual screening rounds. Endpoints in the tree include healthy, local breast cancer, regional breast cancer or distant metastases. Separate mortality rates are described for the three breast cancer states.

Lai et al<sup>6</sup> use two separate Markov chain models that describe only two breast cancer states: regional lymph nodes negative (or tumour size smaller than 2cm), and regional lymph nodes positive (or tumour size larger than 2cm). All states may be either pre-clinical or clinical. Mortality rates are based on the state at diagnosis or detection.

Van Oortmarssen et al<sup>7</sup> present an alternative model of breast cancer screening, based on the general Microsimulation for SCreening ANalysis (MISCAN) approach reported by Habbema et al<sup>8</sup>. The Monte Carlo simulation model consists of a disease and a screening part. The disease part simulates a large number of individual life histories, based on assumptions regarding the epidemiology and natural history of the disease. Then, a specified screening programme is applied to the life histories, which changes some of the histories, constituting the simulated effect of screening.

The breast cancer MISCAN model describes progression through a non-invasive state (dCIS), and three discrete stages of invasive cancer based on the size of the tumour (<10mm, 10-19mm, and >20mm). Tumours in each of these categories may be pre-clinical (i.e. unknown), screen-detected, or clinically diagnosed. The duration of the pre-clinical phase is modelled as being age-specific, based on 10-year age groups. Breast cancer mortality rates are based on the size of tumour at the time of diagnosis or detection, though it is recognised that the inclusion of lymph node status would add to the completeness of the model.

The model presented by Szeto & Devlin<sup>9</sup> is based on the MISCAN breast cancer model, where the main difference is the handling of post-diagnosis/detection survival. Once diagnosed, patients are cured or not, with separate age-dependent survival times applied. Non-cured survival times are referenced from a log-normal model for survival.

The breast cancer screening model developed by Eddy<sup>10</sup> involves a set of formulae describing the probability that a woman will have a cancer detected if screened, or will present with cancer between tests (as a function of risk factors and screening history). If a cancer is detected a second set of formulae calculate how early it was detected as a function of how it was detected and screening history. A third set of formulae calculate the probability of dying from breast cancer at any point in time as a function of the earliness of detection.

Other analytical models,<sup>11</sup> including a recent version,<sup>12</sup> model the natural history of breast cancer as a series of continuous variables. Schwartz, for example, assumes tumours grow at a

continuous rate and that the involvement of lymph nodes is a function of the age of a tumour. However, seven discrete tumour sizes are defined in order that differential sensitivity rates may be applied. The rate of clinical diagnosis (in the absence of screen-detection) is modelled as a function of tumour age, and the mortality rate is modelled as a function of discrete prognostic groups that account for tumour size and lymph node involvement.

Baker<sup>4</sup> lists a similar set of structural assumptions, though the rate of clinical diagnosis is stated to be a function of size and growth rate, as is the breast cancer mortality rate. Potential structural relationships are also discussed, including the likelihood of a non-random element to sensitivity (i.e. the probabilities of detecting cancers at successive screens is not independent), and that the clinical diagnosis rate is also a function of the time since last screen. Both Schwartz(47) and Baker<sup>12</sup> ignore the presence of dCIS.

Finally, Parmigiani<sup>13</sup> presents a stochastic compartment model that describes progression from healthy to detectable pre-clinical cancer to clinical cancer (and to death from any of the preceding states). Transition densities for each possible transition between two states inform the proportion of patients making each transition in each cycle of the model. The estimated transition densities are estimated as continuous functions of the following parameters for each transition:

The duration of the pre-clinical detectable phase (PCDP) is a function of the age at tumour onset, Mortality rates from the clinical state are also a function of age at tumour onset, as well as prognostic variables (tumour size, lymph node involvement and estrogen receptor status),

The severity of the prognostic variables is a function of age at onset and time in the pre-clinical state.

#### Colorectal cancer model structures

The natural history of colorectal cancer includes a non-invasive stage (polyps), which may regress. In addition, adenomatous polyps can become larger over time and may develop into cancer. Hyperplastic polyps do not increase in size and do not become cancerous.

The majority of models developed to analyse the cost-effectiveness of colorectal screening are Markov models, or are assumed to be Markov models (the form of model is not explicated).

Eddy<sup>14</sup> published one of the first Markov models of colorectal cancer screening. This model includes three invasive cancer states (Dukes' A, B and C stages), but it is not clear how the non-invasive stage is modelled. It is likely that three non-invasive states are modelled, which may describe different polyp sizes. Eddy assumes that a set proportion of cancers emanate from

polyps, and specifies a mean duration for the non-invasive phase. The proportion of polyps developing to cancer is fitted to these data and the polyp and cancer incidence rates. Mean durations are also specified for each of the invasive cancer stages.

Ladabaum et al<sup>15</sup> and Shimbo et al<sup>16</sup> use similar model structures and assumptions, with slight variations around the number of invasive cancer stages specified.

Vijan et al<sup>17</sup> also use a Markov model, but incorporate uniform durations through the use of separate year 1 and year 2 local states to represent 2-year duration, and a single regional state to represent 1-year duration, i.e. all persons remain in each state for one annual cycle on their way to the disseminated state. Polyps either remain as polyps or they enter a premalignant dwelling state. A 10-year polyp dwell time is stated, though it is not clear whether this is implemented as a uniform duration, or whether it is transformed into a transition probability and patients leave the dwell state over time.

In addition to defining separate incidence rates for adenomatous polyps and hyperplastic polyps, Vijan et al also define the proportion of multiple polyps and those larger than 1cm, which are used to define surveillance schedules following detection of polyps.

Frazier et al<sup>18</sup> use a Markov model to describe progression between low- and high-risk polyps, and three stages of cancer (local, regional, and distant). However, the two sides of the colon (the distal or proximal colon) are modelled separately because FSIG can only visualise the distal colon. It appears that similar transition rates are applied to polyps originating in either colon. The model does not explicitly assume the proportion of polyps progressing to cancer, though the model structure implies that all cancer originate from polyps. Mean annual transition of the low-risk to high-risk polyps, and from high-risk polyps to early cancer are assumed.

Khandker et al<sup>19</sup> describe a dynamic state transition model. The basic model describes the incidence of hyperplastic and adenomatous polyps. Undetected adenomatous polyps progress to undetected cancer, whilst persons with detected polyps enter a surveillance state following treatment. Undetected cancer progresses to a treatment stage (either through screen-detection or clinical diagnosis). However, more detail is included within this basic structure, for example polyps are described as a function of their size, stage of development and location, whilst invasive cancer incorporates the stage of cancer and the number of years in each stage. Tunnel states are used to apply time-in-state dependent transition probabilities from the polyp state (modelled as 10 states of 2-year duration) and different cancer stages (undiagnosed affected persons could remain in each stage for up to five years).

The Khandker et al model appends decision trees to each of the basic model states to incorporate the extra detail. For example, at the end of each cycle a decision tree is solved for all patients in the adenomatous polyp state, which models the interactions between the screening tests and the characteristics of the adenomatous polyps.

Neilson & Whynes<sup>20</sup> describe a semi-Markov process that builds on the model developed by Wagner<sup>21</sup>. The semi-Markov process samples the state to which each person will move from their current state. Given the state to which the person will next move, a holding time in the current state is sampled from the relevant probability distribution, and the patient remains in the current state for the sampled duration.

The model describes progression from healthy to adenoma to early asymptomatic cancer to late asymptomatic cancer. Age-specific incidence rates are described, but no further transition probabilities are assumed to be age-specific. Cancer may become clinically diagnosed as either early or late.

Two other simulation models were identified. Firstly, the general MISCAN model has been applied to colorectal cancer screening<sup>22,23</sup>. Loeve et al model the progression of separate lesions within individuals. Each individual is assigned a risk index that describes their relative risk of developing polyps. Each polyp is assigned an anatomical site defined in terms of the part of the bowel and a percentage that indicates the localisation within this part. Each lesion may progress through three size-related polyp states (<5mm, 6-9mm, >10mm), and they may transit from any of the polyp states to pre-clinical cancer. Pre-clinical cancer can progress through four stages (1-4) or become clinically diagnosed from any stage, from which point a stage-specific survival time is sampled. A time of death from other causes is sampled for each individual, and this time is applied unless any of the lesions occurring in an individual lead to colorectal cancer death prior to the point of death from other causes.

Pathways between states can depend on age and the anatomical site. For each possible transition between two stages a probability distribution of the dwelling time in the current state is defined. It is also possible to correlate dwelling times in successive states.

Ness et al<sup>24</sup> use DES to describe a basic model that follows the progression of separate adenomas within individuals. An individually assigned risk value is used to adjust the age-dependent polyp incidence rate for each individual. Each polyp is described in terms of its size and location (left, right, sigmoid, high-rectum, low-rectum). Incident polyps are assigned to either a 'fast growing' or 'slow growing' group. All polyps are assumed to progress to cancer, but many will not progress within the natural lifetime of the individual. Polyps move through the same size

categories as defined by Loeve et al<sup>22</sup>, and may progress to cancer from any of the three polyp states. Three cancer states are defined (local, regional and distant).

It is not explicitly stated that the model samples transitions (i.e. the next state to which an individual will move) and then samples a duration prior to the move, though it is implied.

Both the MISCAN and the Ness et al models describe lifetime profiles for a large set of individual and then apply alternative screening schedules to each life history that may alter the course of each person's lifetime events.

All of the above models handle test sensitivity similarly. All studies recognise that the available screening tests have different levels of sensitivity for different sizes of polyp, and for different stages of cancer, though the categories to which alternative sensitivity rates are applied differ between studies. All studies incorporate a non-random aspect to sensitivity, primarily through the reach of FSIG. Lesions out of reach are assumed to be non-randomly false negative, whilst random sensitivity rates are applied to the reachable lesions.

In the above studies, stage at diagnosis is the most common determinant of cancer mortality rates<sup>17;22;24</sup>. Eddy<sup>14</sup> defines sex- and stage-specific mortality rates. Khandker<sup>19</sup> bases mortality rates on stage and the number of years with cancer at diagnosis, though it is not clear how the referenced SEER data incorporates years with cancer into the mortality rate estimates.

Frazier et al<sup>18</sup> state that sex-, race- and stage-specific mortality rates are applied to all cancers regardless of means of detection, or state of detection. It is not clear how colorectal cancer mortality rates can be applied to cancers that remain undiagnosed as the stage at diagnosis will influence treatment effectiveness, and only deaths in which distant cancer is recorded can be defined as colorectal cancer deaths (i.e. deaths in which an undiagnosed person is found to have early stage colorectal cancer at autopsy would be defined as deaths from other causes).

The US Office of Technology Assessment<sup>25</sup> developed what appears to be a Markov model that follows persons from well, to polyp, to early colorectal cancer, to late colorectal cancer (the model structure or type is not defined). A constant proportion of cancers originate as polyps, and a mean duration of the non-invasive phase is specified. If the OTA model is a Markov model, however, it would be unique in the sense that it uses age- and stage-specific mortality rates, which requires the representation of separate states for different age groups to which to apply different age-specific mortality rates.

Wagner et al<sup>21</sup> do not specify the form of model, though it is unlikely to be a Markov model because it uses age- and stage-specific mortality rates, but also because the text refers to the use of uniform transition times (in a Markov model transition times vary within a population

because a mean duration is transformed into cycle-based transition probabilities, such that a proportion of patients exit a state after each cycle).

This model also describes progression from well, to polyp, to early, to late cancer. This model specifies a proportion of early cancers that remain latent over a lifetime. A uniform duration of the non-invasive phase for polyps destined to become cancers is assumed, as is the duration of early stage cancer for those cancers progressing to late cancer.

Gyrd-Hansen et al<sup>26;27</sup> adapt a previously developed statistical model<sup>28</sup>, which requires data describing the numbers of screen-detected and interval cancers, as well as assumptions that sensitivity is independent of disease stage and that the distribution of the pre-clinical detectable phase is exponential. Test sensitivity and average duration of the pre-clinical invasive phase (sojourn time) are estimated using a maximum likelihood estimation method. Estimates of test sensitivity and average sojourn time are combined with age-specific (cancer?) incidence rates in a simulation process to estimate the number of cancers detected at each screening round. It is assumed that 30% of screen-detected cancer patients survive due to early detection, this percentage is assumed constant across different screening intervals. The life years gained from the early detection of cancer are estimated as the sum of the age-specific life expectancies for surviving individuals after adjusting for mean lead time.

The model also accounts for the detection of non-invasive polyps, to which the relative risk of developing cancer is applied. The cancer incidence rates were adjusted for the estimated number of cancers avoided due to the detection of polyps.

Sonnenberg and colleagues<sup>29;30</sup> “tried to reduce the complex natural history of colorectal cancer to few essential states and avoid transition assumptions for which little or no published data existed”. Their Markov model describes patients health states in terms of the screening process: patients are either non-compliant and may progress only to colorectal cancer, their previous screen was negative and they progress to cancer or to the next screen, or they underwent a polypectomy after their previous screen from which they can progress to cancer or back to screening three years later. Colonoscopy, which is undertaken after every positive screen (or as the screen itself), is assumed to reduce colorectal cancer incidence by 75%. Presumably, this risk reduction is applied to persons undergoing colonoscopy. Of the cases that progress to cancer, colonoscopy is assumed to detect the cancer earlier, leading to improved survival.

Comparison of model structures and assumptions

The reviewed models of breast and colorectal cancer screening vary significantly with respect to the level of detail with which they describe the disease process, and its interaction with screening. The majority of the studies model the natural history of a cancer, to which specified screening programmes are applied. The main alternative approach is to model the incidence of cancer and apply trial-based risk reductions to cancer mortality rates to describe the benefits of screening.<sup>3</sup> Sonnenberg and colleagues<sup>29;30</sup> present a variation on the mortality reduction approach, in which screening is assumed to prevent a percentage of cancers, whilst survival rates for early detected cancers are adjusted to account for the observed reduction in colorectal cancer mortality in early detected cancers.

The main disadvantage of the 'mortality reduction' approach is that it requires the application of observed mortality reductions, which reduces the ability to evaluate screening programme configurations (e.g. combinations of screening tests and intervals) for which there are no observed estimates of effect. The identified mortality reduction studies do not provide a clear explanation of their models, which may be due to the need to estimate costs by describing the screening process in the screened population, but then estimating effects on the basis of cancer incidence in the non-screened population.

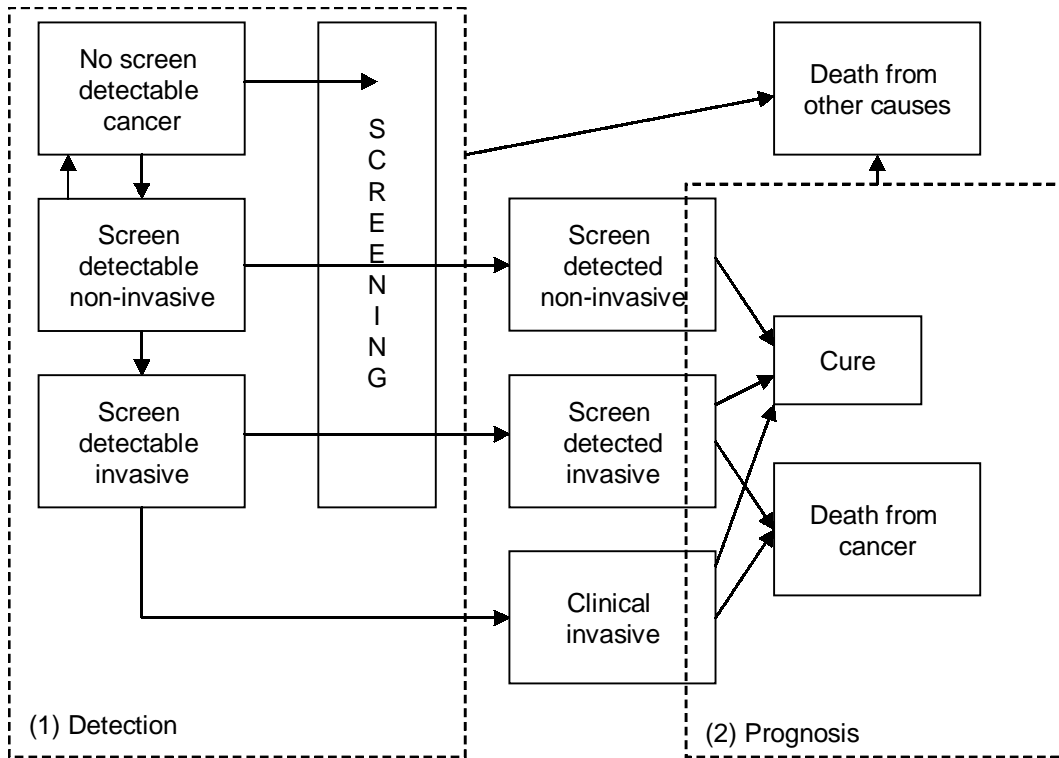
The decision tree models do not describe the natural history of cancer,<sup>4;5</sup> rather they describe alternative stage distributions of cancer at diagnosis, to which separate mortality rates are applied. The obvious disadvantage of the decision tree for modelling screening programmes is that it is difficult to estimate the impact of a series of screening rounds.

Alternative modelling techniques, that model the relevant natural history of a cancer with sufficient accuracy provide the most flexible framework for evaluating a range of screening programmes, particularly for screening programme configurations that have not been evaluated directly.

Van Oortmarssen et al<sup>31</sup> describe a general model of the natural history of cancer and its interaction with screening (

Figure 2). The model defines two cancer states: non-invasive and invasive, though within these states a range of sub-states can be described. The model also allows for the regression of non-invasive disease, such that individuals may re-enter the 'no cancer' state. The state in which cancer is diagnosed (invasive or non-invasive), in combination with the mode of detection for invasive cancer (screen- or clinical-), affects the prognosis of the patient. The general model also assumes that a cancer may be defined as cured. This is a modelling artefact, as it is not possible to define a cured cancer patient following treatment, this state includes all persons who are diagnosed with cancer that do not die from cancer.

**Figure 2** General cancer screening model (reproduced from van Oortmarssen et al, 1995)



This basic structure describes the underlying framework for all of the identified natural history models, though three alternative approaches to representing the structure are identified: Markov models, DES, and complex analytical models.

The Markov model is the standard technique used to model the economic impact of health care interventions over time, and is also the most common approach identified in the review. Events are modelled as transitions from one health state to another. The time horizon covered by the model is split into cycles of equal length. At the end of each cycle a patient may move to a consequent health state, or remain in the same state (unless the current state is a tunnel state). This process of moving between states continues until a patient enters an absorbing state, such as the state 'dead'. Transition probabilities are conditional on the current health state, but they may also vary according to the overall time spent in the model.

Markov models have been shown to be broadly equivalent to more flexible modelling techniques when applied to secondary health care interventions<sup>32</sup>, but the limitations of the Markov approach may be more apparent when applied to the evaluation of screening programmes. The most obvious potential problem is that Markov models grow at an exponential rate as the level of detail

with which the disease and screening process increases. The representation of the natural history is noticeably less complex in the reviewed Markov models, such as in the application of age-specific transition probabilities. Non-Markov models of breast and colorectal cancer screening describe age-specific values for parameters such as the pre-clinical detectable phase (PCDP) and mortality rates from the point of cancer diagnosis, whereas none of the above Markov models incorporate age-specific transition rates within the model (i.e. other than age-specific incidence rates). Reviewing other cancer screening models, only Knox<sup>33</sup> incorporates age-specific durations of the PCDP in a Markov model (for cervical cancer screening), which includes 26 separate health states.

Constant PCDP transition probabilities in the described models may be a reasonable assumption based on data and informed opinions about the natural history of the diseases, though the inclusion of age-specific PCDP rates and mortality rates in other, less restrictive, models indicates that there is some evidence to support their incorporation. Differential age-specific PCDP rates are likely to affect the cost-effectiveness of screening at different ages, for example, if younger age groups have a shorter mean PCDP then, *ceteris paribus*, shorter screening intervals will be more cost-effective at younger ages.

The default application of a cohort-based Markov model may also impose inappropriate restrictions when mean durations are applied to all patients within states that interact with the screening intervention, primarily the health states described within the PCDP. In a Markov model a mean duration is transformed to a transition probability, which describes an exponential distribution of person-level transitions between a PCDP state and a clinical diagnosis state. In a standard evaluation of a treatment intervention, the impact of modelling the wrong distribution of person-level transitions would be limited to a slight error in the discounted outputs from the model (costs and effects). However, in a screening model, the potential error if the real distribution of transitions is not exponential is greater because the screening programme is laid on top of the natural history model, and an inaccurate time-dependent distribution of persons across health states will have some impact on the estimated effectiveness of the screening programme. As an example, if the real distribution of person-level transitions (assuming the same mean value) describes a decreasing transition rate then the Markov approach will artificially favour longer screening intervals.

The representation of person-level distributions of state durations is related to the modelling of non-age-specific PCDP transition probabilities. The latter describes the specification of a mean duration for a specified population (second-order uncertainty), whilst the former concerns the distribution of the mean duration within the population (first-order uncertainty).

The above limitations describe factors that can be overcome by using a more flexible form of decision analytic modelling. DES is an event-orientated modelling approach, whereby the model asks what and when is the next event for every person at the point at which they experience their current event, rather than a Markov model, which asks what events are occurring at regular intervals. DES models use attributes that record relevant elements of each individual's pathway through the model and personal characteristics, which can then influence future pathways, such as duration of the PCDP. By following individuals through the model, it is also possible to assign individual state durations sampled from any form of person-level distribution.

DES models may only be analysed using person-level (first-order) Monte Carlo simulation, which increases the analysis time. Though, the assignment of individual event times facilitates the application of alternative distributional forms to represent parameters such as PCDP duration.

However, to achieve the most accurate representation of reality from the structure of a screening model, incorporating the strongest possible set of assumptions (note: though not necessarily the most accurate model when populated – see input data section below), DES may be superseded by complex analytical models. These models comprise a series of related differential equations that incorporate the assumptions made about the natural history (e.g. that there is a distribution of tumour growth rates), screening (e.g. sensitivity is a function of tumour size), and possibly interactions between disease progression and screening (e.g. the timing of presentation with clinical symptoms is a function of the time since last screen). The equations are solved to produce probability distributions for each output parameter, for example, distributions of tumour sizes for screen-detected and clinically diagnosed cancers, or distributions of survival times for screened and non-screened cohorts.

These more complex analytical models describe input parameters as continuous variables that change smoothly over time, which means that parameters that are functions of other parameters, such as age-specific event rates, are modelled to their most exact specification. An example concerns the description of discrete cancer stages, and the application of stage-specific mortality rates, which does not account for differences between screen-detection and clinical diagnosis within a stage. Applying the same mortality rate to screen-detected and clinical diagnosed patients will bias results against screening because screen-detected cases will, on average, be detected earlier within a stage. The estimation of mortality rates as a continuous function of the stage at diagnosis is more likely to overcome this potential source of bias.

Most of the complex analytical models applied to cancer screening have not incorporated in-depth analyses of costs and effects, i.e. applying separate cost and utility values to different health states along the pathway from model entry to death. Baker,<sup>12</sup> for example, minimises a

cost function to identify the optimal screening programme, in which cost values are applied to each screening test undertaken and to each month of life lost.

Parmigiani,<sup>13</sup> however, applies the outputs from a complex analytical modelling approach to a decision analytic framework, which enables the standard application of cost and utility values to defined health states along person pathways. The analytical outputs are a set of probability distributions describing the proportion of women moving between each pair of states (for which transitions are allowed) over time. The output distributions, for example, the set of distributions describing the person-level distributions of the PCDP for women developing breast cancer at different ages, are combined to calculate the number of women transiting from the pre-clinical to the clinical stage at any point in time.

The comparison of model structures, and the associated assumptions about the disease process, identifies the modelling approach that facilitates the most accurate representation of the process being modelled. However, models are meant to simplify reality and the exclusion of age-specific PCDP rates (and mortality rates), the application of patient-level durations, or the representation of disease progression as a series of continuous variables, does not necessarily mean that less complex models are inappropriate. The obvious difficulty is in establishing the insignificance of such factors when they are not included in the model. In practice, such exclusions are justified (if at all) on the basis of a lack of data or the opinions of experts. More appropriate would be a discussion of the likely direction of broadening the assumptions, i.e. which screening programmes would become more or less cost-effective? If possible, an assessment of the magnitude of the impact should also be included. If a reasonable case can be made that a more complex representation of reality would not alter the conclusions drawn from the analysis, then a less complex model may suffice.

The appropriate level of model complexity, and the reasoning behind simplifying assumptions should be an integral part of the initial development of the model structure and choice of modelling technique. Such assumptions are likely made before the full available evidence has been reviewed and so modelling assumptions may change over the course of a study. Even if modelling assumptions do not alter, available evidence around the areas in which simplifying assumptions have been made should be collated to inform discussion around the appropriateness of the model.

As noted above, the limitations of the more simple decision analytic models may be more exposed when applied to the evaluation of screening programmes due to the level of interaction between the natural history of a disease and the screening programme. Therefore, the additional

flexibility provided by the more complex modelling techniques may be more likely to result in alternative conclusions being drawn from alternative modelling approaches.

#### Comparison of broad parameterisation approaches

The previous section describes alternative model techniques that facilitate more or less accurate representations of the natural history of cancer, and its interaction with screening programmes. However, the relevance of the outputs from the most comprehensive model structure will be limited if the values of the input parameters are based on less than optimal data sources, and parameter estimation methods. This section describes three broad approaches to model parameterisation that were identified during the review of cancer screening models.

The first approach describes a process of fitting a wholly unpopulated model to all available data simultaneously, as demonstrated by Baker,<sup>12</sup> who uses maximum likelihood estimation to fit the model. The likelihood function “is a complicated function of data and model parameters that for this problem requires numerical evaluation of double and triple integrals” [pp103-4]. Datasets describing five types of data are used to populate the model, including the tumour size for screen-detected cancers, the timing and tumour size for interval cancers, survival times for screen- and clinically-detected cancers, pre-screening age-specific cancer incidence, and all cause mortality rates. The former three types of data are derived from the same source.

An objective measure of the best fitting parameterization is possible – minimising the Akaike Information Criterion (AIC). Baker has published a separate paper that describes a form of sensitivity analysis for models fitted to data by statistical methods, which uses the covariance matrix of the fitted model parameters.<sup>34</sup>

The fully fitted parameterization approach requires separate datasets describing cancer outcomes in the presence and absence of screening for the same population. The comparability of these datasets in the Baker paper is good because the non-screened cancer incidence rates from 1987 are compared with screened incidence rates from 1988-1990. However, the data describing non-screened and screened incidence rates will become less comparable as the distance between the non-screened and screened cohorts widens. It is not clear how the statistical methods described could compensate for this divergence.

An advantage of fully fitted parameterisation approach is that it is based solely on data from a specific population, which makes the results of the model particularly relevant to that population. Baker uses data collected at three screening centres in the north west of England. In the UK, screening policy is defined at a national level, which raises the issue of the generalisability of a

model based on three screening centres in one area of the country. One approach to testing the generalisability of the model would be to fit the model to data derived from alternative populations and then to compare input parameter estimates, especially for input parameters that are not likely to vary across populations, for example, age-specific tumour growth rates.

A related criticism of the fully fitted approach is that it uses only a small amount of the data available to inform cancer screening models. By fitting the model to a relatively small observed dataset, the model is effectively ignoring the vast amount of research that has been conducted into different aspects of the natural history of the disease and the screening test characteristics. However, if the model could be fitted to relevant data covering the whole of the UK, such criticisms may be of less concern.

The second identified parameterisation approach is to populate parameters directly for which specific data (of an implicit level of quality) are available. Explicit or implicit assumptions are made about the distributional form of the unpopulated parameters, which are then estimated simultaneously against observed data. This multiple fitted parameter approach involves simulating a wide range of combinations of parameter estimates for the unknown parameters, and estimating the goodness-of-fit between the observed and simulated results for each combination.

The MISCAN breast cancer model, for example, treats as unobservable the duration of the PCDP (split into two states) and the sensitivity of the screening test in these two PCDP states.<sup>7</sup> The best fitting point estimates for these parameters are presented, as well as an 'area of combinations' of total PCDP duration and average sensitivity that are in agreement with the results of the observed study. It is not clear how the area of combinations is defined as the authors state that the fit of the point estimates for interval cancers by age and interval cancers by time since last negative screen is outside the conventional threshold ( $p > 0.05$ ), such that the area of combinations cannot comprise all combinations that produce  $p < 0.05$  for all data.

The objective of the fitting process is to define a best fitting set of parameters values that may then be used to test the impact of alternative screening programmes. Problems in defining point estimates may occur if a wide range of parameter value combinations provide a similar goodness-of-fit – should the baseline analysis use the combination that maximises the goodness of fit, or the midpoint of the area of combinations that are within a defined threshold of goodness-of-fit? Church suggests that as more than one combination of parameter estimates could result in a similar goodness-of-fit, a full analysis of the uncertainty around the combinations of the fitted parameters is necessary.<sup>35</sup> Presumably, second-order probabilistic sensitivity analyses could be

undertaken using probability distributions based on the 'area of combinations', though this approach does not appear to have been applied to date.

The multiple fitted parameter approach suffers from the same generalisability problem as the fully fitted parameterisation approach, if the dataset to which the model parameters are fitted relates to only a small subset of the relevant population. In the original description of the MISCAN approach, Habbema et al stated that ideally, the goodness-of-fit would be tested against a series of implemented screening programmes, which would allow the best combination of assumed data inputs to be chosen.<sup>8</sup> Van Oortmarssen et al did implement this approach using the breast cancer MISCAN model, in which model predictions are compared to observed data from two alternative Dutch screening programmes.<sup>36</sup> Using parameter estimates fitted to a previous trial (the HIP trial) resulted in a poor fit, so input parameters were varied systematically to obtain 'an adequate overall fit of almost all screening results' from the two areas. Possible reasons for the lack of fit in some areas are presented (relating to the implementation of the screening programme).

This process of cross-validation using more than one set of observed data requires an appraisal of the factors that are most likely to stay constant across different populations, for example, genetic factors, such as prevalence of the breast cancer gene, and environmental factors, such as smoking rates, will affect incidence rates, but how do such factors affect growth rates?

The multiple applications of the MISCAN model provide examples of the multiple parameter fitted approach,<sup>7;37;38</sup> though the colorectal cancer MISCAN model appears to fit only one parameter – the incidence of non-progressing adenomas.<sup>22</sup> The DES model presented by Ness et al fits multiple input parameters,<sup>24</sup> whilst most of the identified Markov models estimate all input parameters individually. The less complex representation of the PCDP in most Markov models requires simpler estimates of parameters such as PCDP duration and test sensitivity, i.e. a single value for each parameter, rather than multiple values for each, as well as probability distribution parameters for each defined PCDP state. Thus, the data requirements to populate all model parameters directly are reduced.

As alluded to in the previous paragraph, the third broad approach to model parameterisation is to estimate all input parameters directly. This approach does require some basis for the estimation of the more difficult model parameters, though various research methods have been developed to estimate parameters such as PCDP duration<sup>39-41</sup>, as well as test sensitivity<sup>42;43</sup>. Though the majority of models that use this approach describe relatively simple model structures and assumptions, implying that the direct estimation approach does not facilitate complex representations of these parameters, some of the developed methods do describe these parameters in detail, for example, Ashih estimates mammography sensitivity as a logistic function

of tumour size and age.<sup>44</sup> Parmigiani presents an application of the direct estimation approach in which complex, continuous, parameter values are represented.<sup>13</sup>

The direct estimation of parameters allows for the more comprehensive use of the data available to inform these parameters, though consideration must be given to impact of local factors and variability across locations. If possible, observable local factors should be accounted for in the estimation process.

The process of validation is also crucial. Validation would preferably be undertaken against data derived from the full population for whom the model is to predict cost-effectiveness, normally the national population. Such data may include age-specific screen- and clinical diagnosis, and cancer mortality rates.

Initial model specifications may result in poor representations of observed data, which requires “reconsideration” of original parameter estimates. However, a key advantage of the direct estimation approach is that it provides a grounded basis for parameter estimates and iterative process for re-estimation.

## Discussion

This paper has described the results of a review of modelling studies evaluating screening programmes for breast cancer and colorectal cancer. A summary of the applied modelling approaches provides some background to a comparison of alternative model structures and associated assumptions about the disease process and its interaction with screening, and to a comparison of the broad approaches to populating the screening models.

Three broad modelling approaches are identified: cohort Markov models, patient-level DES models, and complex analytical models. DES allows a more detailed representation than the Markov, and the complex analytical approach provides more detail than DES. The description of the specific issues that could be modelled in more detail indicate that there may be greater justification for the use of more complex modelling techniques in the evaluation of screening programmes.

Other issues around the model structure include the level of detail used to model post-diagnosis survival. The vast majority of model use observational estimates of survival to which mean cost and utility values are applied. Only one model identified across all cancer screening models incorporates a treatment model from the point of diagnosis.<sup>45</sup> The more detailed treatment model approach allows for the use of predictive models of survival that are based on current (or future)

treatment protocols, whilst observational studies will likely be based on old management pathways that may not reflect current practice.

There is also the question of whether to model the lifetime models<sup>46</sup> versus cross-sectional models over time.<sup>47</sup> Lifetime models describe the experience of a single birth cohort over their lifetime, whilst the alternative evaluates screening over a specified time period over which time women of all birth cohorts attend. Cross-sectional models allows the use of birth cohort-specific incidence rates, and incorporates the impact of women starting screening at different ages. The latter point highlights the estimation of cost-effectiveness when screening has reached a steady-state, and in the period prior to steady-state. The presentation of results from a cross-sectional model should emphasis the distinction between these two states.

Comparing the alternative parameterisation approaches in the context of evaluating screening programmes at the national level (as is the case in the UK), there is a trade-off between estimating parameter values from a localised dataset and then generalising the model to a wider population, and estimating parameter values using all relevant data sources (for example, a meta-analysis of sensitivity studies) and applying these estimates to a specific population.

The two most difficult parameters are the duration of the PCDP and the screening test sensitivity. The duration of PCDP may be hypothesised to vary only slightly across different populations, such that direct estimates using all available data may be more appropriate. Test sensitivity may be subject to greater localisation, such that local data (if felt to be sufficiently generalisable) may provide more relevant parameter estimates. However, a key output from the evaluative process may be to inform optimal processes of delivery and so it may be more appropriate to define parameters such as test sensitivity (and the related cost of screening) on the basis of optimal strategies, to which applied screening programmes should aspire. In the latter case, it would make sense to estimate sensitivity directly, rather than to fit the values to localised data.

Another issue is whether it is more appropriate to populate these key input parameters simultaneously or independently. There is obviously a relationship between the duration (and person-level distribution) of the PCDP and test sensitivity which is automatically captured when the two parameters are fitted simultaneously. However, independent estimates of sensitivity may also incorporate such a relationship through the specification of sensitivity as a function of tumour size or stage.

Ideally, the modelling process would incorporate a process of validation that, through an iterative process, would lead to the derivation of similar first-order parameter estimates using both approaches, i.e. the model would be validated (or fitted) against a series of relevant datasets (e.g. alternative local datasets from within the decision-making domain).

## References

1. Wilson, J. M. C. and Junger, Y. G. Principles and practice of screening for disease. 1968. Geneva, World Health Organisation.
2. Lindfors KK, Rosenquist J. The Cost-effectiveness of Mammographic Screening Strategies. *JAMA* 1995;**274(11)**:881-4.
3. Salzman P, Kerlikowske K, Phillips K. Cost-effectiveness of extending screening mammography guidelines to include women 40 to 49 years of age. *Annals of Internal Medicine* 1997;**127(11)**:955-65.
4. Allen MW, Hendi P, Bassett L, Phelps ME, Gambhir SS. A study on the cost effectiveness of sestamibi scintimammography for screening women with dense breasts for breast cancer. *Breast Cancer Research & Treatment* 1999;**55(3)**:243-58.
5. Mandelblatt, J., Wheat, M. E., Monane, M., Moshief, R. D., Hollenberg, J. P., and Tang, J. Breast cancer screening for elderly women with and without comorbid conditions. A decision analysis model. *Annals of Internal Medicine* 116, 722-730. 1992.
6. Lai, M. S. Efficacy of breast-cancer screening for female relatives of breast-cancer-index cases: Taiwan multicentre cancer screening (TAMCAS). *International Journal of Cancer* 78:21-6. 1998.
7. van Oortmarsen, G. J., Habbema, J. D. F., Lubbe, J. Th. N, and van der Mass, P. J. A model-based analysis of the HIP project for breast cancer screening. *International Journal of Cancer* 46, 207-213. 1990.
8. Habbema, J. D. F., van Oortmarsen, G. J., Lubbe, J. Th. N, and van der Mass, P. J. The MISCAN simulation program for the evaluation of screening for disease. *Computer Methods and Programs in Biomedicine* 20, 79-93. 1984.
9. Szeto KL, Devlin NJ. The cost-effectiveness of mammography screening: evidence from a microsimulation model for New Zealand. [Review] [39 refs]. *Health Policy* 1996;**38**:101-15.
10. Eddy, D. M. Screening for breast cancer. *Annals of Internal Medicine* 111, 389-399. 1990.
11. Shwartz M. A mathematical model used to analyze breast cancer screening strategies. *Operations Research* 1978;**26**:937-55.
12. Baker RD. Use of a mathematical model to evaluate breast cancer screening policy. *Health Care Management Science* 1998;**1**:101-13.

13. Parmigiani, G. Decision Models for Screening in Breast Cancer. Available from: <http://www.biostat.jhsph.edu/biostat/event/seminar/semparmigiani.shtml> [Accessed 30 June 2003] . 2003. John Hopkins University.
14. Eddy, D. M. Screening for colorectal cancer. *Annals of Internal Medicine* 113, 373-384. 1990.
15. Ladabaum U, Chopra CL, Huang G, Scheiman JM, Chernew ME, Fendrick A. Aspirin as an adjunct to screening for prevention of sporadic colorectal cancer: a cost-effectiveness analysis. *Annals of Internal Medicine* 2001;**135(9)**:769-81.
16. Shimbo, T. Cost-effectiveness analysis of strategies for colorectal cancer screening in Japan. *International Journal of Technology Assessment in Health Care* 10(3):359-375. 1994.
17. Vijan S, Hwang EW, Hofer TP, Hayward RA. Which colon cancer screening test: a comparison of costs, effectiveness, and compliance. *American Journal of Medicine* 2001;**111(8)**:593-601.
18. Frazier, A. L. Cost-effectiveness of screening for colorectal cancer in the general population. *Journal of the American Medical Association* 284(15):1954-1961. 2000.
19. Khandker RK, Dulski JD, Kilpatrick JB, Ellis RP, Mitchell JB. A decision model and cost-effectiveness analysis of colorectal cancer screening and surveillance guidelines for average-risk adults. *International Journal of Technology Assessment in Health Care* 2000;**16**:799-810.
20. Neilson AR, Whynes DK. Cost-effectiveness of screening for colorectal cancer: a simulation model. *IMA Journal of Mathematics Applied in Medicine & Biology* 1995;**12(3-4)**:355-67.
21. Wagner, J. Cost effectiveness of colorectal cancer: screening in the elderly. *Annals of Internal Medicine* 115:807-817. 1991.
22. Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JD. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. *Computers & Biomedical Research* 1999;**32**:13-33.
23. Loeve, F. Endoscopic colorectal cancer screening: a cost-saving analysis. *Journal of the National Cancer Institute* 92(7), 557-563. 2000.
24. Ness RM, Holmes AM, Klein R, Dittus R. Cost-utility of one-time colonoscopic screening for colorectal cancer at various ages. *American Journal of Gastroenterology* 2000;**95(7)**:1800-11.

25. US Office of Technology Assessment. The cost-effectiveness of colorectal cancer screening in average-risk adults. OTA-BP-H-146. 1995.
26. Gyrd-Hansen, D., Sogaard, J., and Kronborg, O. Colorectal cancer screening: efficiency and effectiveness. *Health Economics* 7, 9-20.
27. Gyrd-Hansen, D., Sogaard, J., and Kronborg, O. Analysis of screening data: Colorectal cancer. *International Journal of Epidemiology* 26(6), 1172-1181. 1997.
28. Day, N. E. and Walter, S. D. Simplified models for screening for chronic disease: estimation procedures for mass screening programmes. *Biometrics* 40, 1-14. 1984.
29. Sonnenberg A, Delco F, Inadomi JM. Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Annals of Internal Medicine* 2000;**133(8)**:647-9.
30. Sonnenberg A, Delco F. Cost-effectiveness of a single colonoscopy in screening for colorectal cancer. *Archives of Internal Medicine* 2002;**162**:163-8.
31. van Oortmarssen, G. J., Boer, R., and Habbema, J. D. F. Modelling issues in cancer screening. *Statistical Methods in Medical Research* 4, 33-54. 1995.
32. Karnon, J. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Economics* 12(10), 837-848. 2003.
33. Knox EG. A simulation system for screening procedures. In: The Future and Present Initiatives. In McLachlan, ed. *Problems and Progress in Medical Care IX*, p 19. Oxford: Nuffield Provincial Hospitals Trust, 1973.
34. Baker RD. Sensitivity analysis for healthcare models fitted to data by statistical methods. *Health Care Management Science* 2002;**5**:275-81.
35. Church, T. R. Simulated screening for prostate cancer: the useful model. *Journal of the National Cancer Institute* 95(12), 838-839. 2003.
36. van Oortmarssen, G. J., Habbema, J. D., van der Mass, P. J., de Koning, H., Collette, H. J. A., Verbeek, A. L. M., Geerts, A. T., and Lubbe, K. T. N. A model for breast cancer screening. *Cancer* 66, 1601-1612. 1990.
37. Habbema, J. D. F., van Oortmarssen, G. J., Lubbe, J. Th. N, and van der Mass, P. J. Model building on the basis of Dutch cervical cancer screening data. *Maturitas* 7, 11-20. 1985.
38. Draisma, G., Boer, R., Otto, S. J., van der Crujisen, I. W., Damhuis, R. A. M, Schroder, F. H., and de Koning, H. Lead times and overdetection due to prostate-specific antigen screening:

estimates from the European randomised study of screening for prostate cancer. *Journal of the National Cancer Institute* 95, 868-878. 2003.

39. Yakovlev AY, Tsodikov AD. STOCHASTIC MODELS OF TUMOR LATENCY AND THEIR BIOSTATISTICAL APPLICATIONS. World Scientific, 1996.

40. Walter, S. D. and Day, N. E. Estimation of the duration of a pre-clinical disease state using screening data. *American Journal of Epidemiology* 118, 865-886. 1983.

41. Peer, P., van Dijck, J. A. A. M., Hendricks, J., Holland, R., and Verbeek, A. L. M. Age-dependent growth rate of primary breast cancer. *Cancer* 71, 3547-3551. 1993.

42. Smith JE, Winkler RL. Casey's problem: Interpreting and evaluating a new test. *Interfaces* 1999;29:63-76.

43. Launois R. Cost-effectiveness analysis of strategies for screening prostatic cancer. [Review] [44 refs]. *Developments in Health Economics & Public Policy* 1992;1:81-108.

44. Ashih, H. Modeling natural history of pre-clinical breast cancer. PH.D.Thesis, ISDS, Duke University . 1998.

45. Coley CM, Barry D, Fleming C. Early detection of prostate cancer: Part II: Estimating the risks, benefits, and costs. *Annals of Internal Medicine* 1997;126:468-79.

46. Jenkins, D., Sherlaw-Johnson, C., and Gallivan, S. Can papilloma virus testing be used to improve cervical cancer screening? *International Journal of Cancer* 65, 768-773. 1996.

47. Parkin, D. M. A computer simulation model for the practical planning of cervical cancer screening programmes. *British Journal of Cancer* 51, 551-568. 1985.