Frailty and Mixture Models in Cancer Screening Evaluation

Ming-Fang Yen

Cancer Research UK Centre for Epidemiology, Mathematics and Statistics

Department of Statistical Science, University College London

October 2004

Dissertation submitted to the University of London

In fulfilments of requirements for the degree of Doctor of Philosophy

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The prevalence of screen-detected premalignancies is too large for it to be feasible that all can progress to carcinoma at the same average rate, unless that rate is very low indeed. There are likely to be frailties in the rates of progression. Failure to take heterogeneity into account will lead to biased estimates and could result in inappropriate screening policy. Approaches to investigation of heterogeneity in the propensity for screen-detected disease to progress comprise the main objectives of this project.

We used Markov models with constant hazard rates in sequence throughout the process of disease natural history within subjects, with heterogeneity terms by means of (1) frailty models for continuous heterogeneity, (2) mover-stayer models for dichotomous heterogeneity (in both cases for progression between sequential homogeneous models), and (3) latent variables and states to estimate the parameters of progressive disease natural history in the presence of unobserved factors. Approaches had to be developed to address problems of tractability and estimation. For example, in the presence of frailty, solution of the Kolmogorov equations by routine matrix algebra is no longer possible. Heterogeneous models, both discrete and continuous, were found to be tractable, and estimation was possible for a variety of designs and data structures. Such models illuminated various issues in real screening applications.

Quantifying heterogeneity of potential progress of disease is of potential importance to the screening process. There are trade-offs between model complexity, identifiability and data availability, but there are clear examples, such as that of cervical screening, where a heterogeneous model improves model fit and gives more realistic estimates than a homogenous.

Acknowledgement

I would like to express my deeply gratitude to my supervisor Professor Stephen Duffy from Cancer Research UK Centre for Epidemiology, Mathematics and Statistics whose help, stimulating suggestions, wide knowledge of cancer screening, encouragement and patience helped me in all the time of research for and writing of this thesis.

I would like to thank Professor Tom Fearn and Professor Stephen Senn from Department of Statistical Science, University College London for their kindly help and concerns. Many thanks are given to Professor Eugenio Paci and Dr. Bianca De Stavola for their valuable comments on the thesis during my viva.

I am deeply indebted to Professor Tony Hsiu-Hsi Chen from Institute of Preventive Medicine, National Taiwan University, Taipei, Taiwan, who introduced me to the field of stochastic process and the application on cancer screening. His logical thinking of applying quantitative methods on biomedical data and enthusiasm on research affect me so much on my career.

I warmly thank Professor Laszlo Tabar from the Department of Mammography, Falun Central Hospital, Sweden. His extensive knowledge, focus, and confidence on, and his concerns about the effect for women from mammographic screening provided a paragon of research.

My sincere thanks are due to colleagues and friends in London and all friends from Taiwan for their concerns and supports.

I would like to give my special thanks to my family whose patient love enabled me to complete this work.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Basic aims

Often in cancer screening, the target of the screening test is a preinvasive or even premalignant state, for example cervical intraepithelial neoplasia or polyps in the colon. The prevalence of these premalignant conditions is too large for it to be feasible that all can progress to carcinoma at the same average rate, unless that rate is very low indeed. Thus there is likely to be frailty in the rate of progression due to unobserved covariates, and possibly even mover-stayer population mixtures in which some lesions simply cannot progress at all. Another application area is in assessment of cancer rates after a negative screen for cancer. Some cancers occurring early after a negative screen are tumours which were missed at screening and are therefore still at an early stage. Others are at an early stage because they have had little time to progress. A third group may be at a more advanced stage if their early appearance after a negative screen is due to their being more aggressive cancers with more rapid progression. Clearly, quantifying such heterogeneity of potential progression of disease is of some importance to the screening process.

In the past, frailty models have been focused particularly on survival analysis. Stratified and regression analyses incorporating known risk factors are widely applied to interpret the heterogeneity within and among populations. These methods, however, cannot deal with the problem when the risk factors are inaccessible for specific types of data, unable to be measured or even unknown. The failure to take heterogeneity into account will lead to biased results. When severe frailty exists, the population hazard will increase to a peak then decline with time due to the effect of selection and this could be incorrectly interpreted as the central tendency of individual hazard. Or even worse, when there is a large proportion of

non-susceptible cases that reduce the average population hazard, it is wrong to apply the resulting small hazard to all individuals, as the aforementioned situation.

The thesis will aim at the development and application of simple heterogeneity and frailty models to address the issues mentioned above, particularly in relation to progression of preclinical conditions, and to draw some conclusions about cancer screening strategy from the application of these methods to screening data.

1.2 Basic definition and applications of frailty and heterogeneity models

1.2.1 Frailty models – mathematical development for failure time data

Since Vaupel and colleagues[1] introduced the concept of frailty to model the different susceptibilities in a population, frailty models have been extensively applied to the time to failure event analysis.

A frailty model is a random effects model for time variables, where the random effect (the frailty) has a multiplicative effect on the hazard. The hazard for a person with time-independent frailty z is assumed to be of the form

$$\lambda(t,z) = z\lambda(t). \qquad [1.1]$$

The unity frailty mean is assumed when the scale parameter is included in $\lambda(t)$. As z is not observed for each individual, we consider it random and integrate it out. So far a gamma distributed frailty is the most common form for its mathematical convenience, with the closure property of which the distribution among survivors is also gamma with the original shape parameter and the distribution among deaths at a given time is also gamma but with different shape and scale parameters. Other distributions used to describe the frailty include a two point distribution, the uniform distribution, the Weibull distribution, and the log normal distribution. All the nonnegative exponential families including z as canonical statistic have been proved to share the same closure property.[2] The derivation is simply described as follows.

Let $\Lambda(t)$ denote the integrated baseline hazard, $\Lambda(t) = \int_0^t \lambda(u)du$. Combining with expression [1.1], it is easy to show that the survival function conditional on a given frailty is

$$S(t \mid z) = \exp\{-z\Lambda(t)\} \qquad [1.2]$$

Therefore, the unconditional survival function is

$$\begin{aligned} S(t) &= E_z[S(t \mid z)] \\ &= \int \exp\{-z\Lambda(t)\} f(z)dz \qquad [1.3] \\ &= L\{\Lambda(t)\}, \end{aligned}$$

where L(s) denotes the Laplace transform of s. For the nonnegative exponential family, expressed as $P(\delta, \theta)$, with shape parameter δ and scale parameter θ having the probability density function as

$$f(z) = \frac{z^{\delta} e^{-\theta z} m(z)}{\phi(\delta, \theta)}, \qquad [1.4]$$

where $m(\cdot)$ is any function of z not involving the distribution parameters, and $\phi(\cdot)$ is any function of shape and scale parameters not involving the random variable z.

The Laplace transformation for the nonnegative exponential family is

$$L(s) = \frac{\phi(\delta, \theta + s)}{\phi(\delta, \theta)}. \qquad [1.5]$$

According to expression [1.3] the unconditional survival function is generalised as

$$S(t)=\frac{\phi(\delta,\theta+\Lambda(t))}{\phi(\delta,\theta)}. \qquad [1.6]$$

A simple parametric frailty model would be one when time to event is distributed as exponential conditional on the rate $\lambda(z)=z\cdot\lambda$, and then $\lambda(z)$ varies among subjects with a gamma distribution, say the frailty has the density as $f(z)=\frac{\theta^{\delta}z^{\delta-1}e^{-\theta z}}{\Gamma(\delta)}$. According to expression [1.4], it has $m(z)=z^{-1}$ and $\phi(\delta,\theta)=\Gamma(\delta)/\theta^{\delta}$.

The unconditional survival function is then obtained from expression [1.6] as

$$\begin{aligned} S(t) &= \frac{\Gamma(\delta)/(\theta+\lambda t)^{\delta}}{\Gamma(\delta)/\theta^{\delta}} \\ &= \theta^{\delta}(\theta+\lambda t)^{-\delta} \\ &= \left(\frac{\theta}{\lambda}\right)^{\delta}\left(t+\frac{\theta}{\lambda}\right)^{-\delta}, \end{aligned}$$

with the unconditional survival time distributed as Pareto.

The density of frailty among deaths at a given time is

$$f(z\mid T=t)=\frac{f(T=t\mid z)f(z)}{f(T=t)}.$$

From expressions [1.2], [1.3] and the equation that $f(.)=-dS(t)/dt$, the above can be expressed as

$$=\frac{z\exp\{-z\Lambda(t)\}f(z)}{-L'(\Lambda(t))}.$$

When δ is not a function of θ, it can be shown that $\phi(\delta+1,\theta)=-d\phi(\delta,\theta)/d\theta$. For exponential families, the density among deaths is then

$$\begin{aligned} &= \frac{z^{\delta+1}e^{-(\theta+\Lambda(t))z}m(z)}{\phi(\delta+1,\theta+\Lambda(t))} \\ &= P(\delta+1,\theta+\Lambda(t)). \end{aligned} \qquad [1.7]$$

The density of frailty among survivors in time t is

$$f(z \mid T \geq t) = \frac{f(T \geq t \mid z)f(z)}{f(T \geq t)} = \frac{S(t \mid z)f(z)}{S(t)}$$

$$= \frac{\exp\{-z\Lambda(t)\}f(z)}{L(\Lambda(t))}.$$

For exponential families, it can be expressed as

$$\begin{aligned} &\frac{z^{\delta}e^{-(\theta+\Lambda(t))z}m(z)}{\phi(\delta,\theta+\Lambda(t))} \\ &= P(\delta,\theta+\Lambda(t)), \end{aligned} \qquad [1.8]$$

the same distribution of frailty with the original shape parameter.

Thus, the mean of this distribution, i.e. the mean frailty among survivors, is

$$E_{t+}(z) = \int z f(z \mid T \geq t) dz = \int \frac{z \exp\{-z\Lambda(t)\} f(z)}{L(\Lambda(t))} dz$$

$$= -\frac{L'(\Lambda(t))}{L(\Lambda(t))} \int \frac{z \exp\{-z\Lambda(t)\} f(z)}{-L'(\Lambda(t))} dz \qquad [1.9]$$

$$= -\frac{L'(\Lambda(t))}{L(\Lambda(t))},$$

for exponential families,

$$= \frac{\phi(\delta+1, \theta+\Lambda(t))}{\phi(\delta, \theta+\Lambda(t))}$$

In general, the integrated population hazard is

$$H(t) = -\log\{S(t)\}$$

$$= -\log\{L(\Lambda(t))\},$$

for exponential families,

$$= -\log\left(\frac{\phi(\delta, \theta+\Lambda(t))}{\phi(\delta, \theta)}\right). \qquad [1.10]$$

The population intensity is

$$h(t) = H'(t)$$

$$= -\frac{L'(\Lambda(t))}{L(\Lambda(t))} \cdot \lambda(t) = E_{t+} \cdot \lambda(t), \qquad [1.11]$$

which demonstrates that the population density is the average hazard among the survivors.

Hougaard also demonstrated that the choice of distribution was crucial. With Gamma distributed frailty the relative heterogeneity is constant, however, with inverse Gaussian distribution the surviving population becomes homogeneous with time.[2] Hougaard generalised the above distributions in addition to the stable distributions on the positive numbers and the degenerate distribution into a three-parameter family of distribution, $P(\alpha,\delta,\gamma)$, on the positive numbers.[3;4] The Laplace transform of the distribution is

$$L(s)=\exp\left\{\frac{\alpha}{(1-\alpha)\delta}\left[1-\left(1+\frac{\delta\gamma}{\alpha}s\right)^{1-\alpha}\right]\right\}, \qquad [1.12]$$

where $0\le\alpha\le 1$, δ (the squared coefficient of variation) ≥ 0, and γ (the expectation of frailty) >0. If α or δ equals 0, the frailty distribution is degenerate at γ, implying no heterogeneity. When $\alpha=1$, it gives a special case of gamma distribution with Laplace transform as

$$L(s)=\left\{\frac{1}{1+\delta\gamma s}\right\}^{1/\delta}. \qquad [1.13]$$

By extending the range of α to be greater than 1, Aalen[3;5] included a compound Poisson distribution with a nonsusceptible subgroup of positive probability

$$P(Z=0)=\exp\left\{-\frac{\alpha}{\delta(\alpha-1)}\right\} \qquad [1.14]$$

According to the expressions [1.3] and [1.12], the population survival function for this

three-parameter family is

$$S(t)=\exp\left\{\frac{\alpha}{(1-\alpha)\delta}\left[1-\left(1+\frac{\delta\gamma}{\alpha}\Lambda(t)\right)^{1-\alpha}\right]\right\} \qquad \text{if } \alpha\neq 1,\alpha>0, \qquad [1.15]$$

$$S(t)=\left\{\frac{1}{1+\delta\gamma\Lambda(t)}\right\}^{1/\delta} \qquad \text{if } \alpha=1, \qquad [1.16]$$

with the corresponding population intensity

$$h(t)=-\frac{d}{dt}\ln S(t)=\frac{r\lambda(t)}{\left\{1+\alpha^{-1}\delta\gamma\Lambda(t)\right\}^{\alpha}} \qquad \text{for } \alpha>0. \qquad [1.17]$$

The frailty model has been extended to consider the effect of covariates combined with the Cox regression model by specifying the individual hazard as $z\exp(\beta' X)\lambda(t)$, where X is the vector of known covariates and β is the regression coefficient vector. Maximum likelihood can be used when the underlying intensity can be described by finite parameters.[6] The EM algorithm can be used for the semi-parametric case, allowing the underlying hazard to be distribution free.[7] A Bayesian approach may also be used, carrying out with Monte Carlo simulation.[8;9]

In general, frailty models can be classified into two categories: univariate and multivariate failure time analysis according to the source of the variability in time to the specific event.[10] The univariate (independent) failure time analysis is for the case when the random effect is an individual variable, while the multivariate (dependent) failure time analysis is used when the random effect is a variable common to several records and deals with unobserved correlation

via latent variables (frailties) from the same clusters.

1.2.2 Frailty models – applications and practical problems

Applications of univariate frailty models may be seen in incidence and progression of disease. Various frailty models with individual hazard increasing with time have been fitted for patients with diabetes mellitus to data on the incidence of diabetic nephropathy that increase until 20 years duration of diabetes and later decreases.[11] Although a good fit is seen in some models, the estimated degree of heterogeneity and the effect of covariates on the hazard are dependent on the choice of models.[11]

Aalen et al[12] performed a compound Poisson distributed frailty model of the selection phenomenon to interpret the age-specific incidence of testis cancer, which increased until the age group 30-34 years then declined with age in Norway males. In their model, the baseline hazard is assumed as Weibull distributed, hence an increasing hazard of testis cancer incidence within individuals. The model including the year of birth showed how the size of frail group changed over different birth cohorts.

Mortality after myocardial infarction is high and the complication is frequent during the first days. Hougaard[13] used frailty models to allow for inter-individual heterogeneity as a possible explanation of the high hazard rate in early days. Alternatively, the decreasing hazard can be explained as a general decrease in risk for all patients with time since the onset of myocardial infarction. Unlike the incidence of carcinoma or other chronic diseases exemplified above in which the theory of different susceptibility is preferred because of its biological plausibility, it is difficult in this application to prove which explanation is correct; hence the results should be interpreted with caution.

Another application of the frailty model is for the unmeasured host response after vaccination. The true vaccine efficacy can be distorted when there is failure to take the heterogeneity in host susceptibility in both vaccinated and unvaccinated people into account. Longini et al[14] and Halloran et al[15] applied a frailty mixture model for estimating the efficacy of a measles vaccine allowing a population fraction with nonsusceptibility. Boily et al[16] and Desai et al[17] addressed the necessity of studying the behaviour of vaccine efficacy measures under heterogeneous conditions of population in the planning of phase III HIV vaccine efficacy trials, and proposed a new efficacy measure based on log-spline hazard regression to give valid estimation across different modes of vaccine action and in the presence of frailty effects.

Over-parameterisation and identifiability problems can arise in univariate survival analysis where there is only one endpoint variable per individual. Introducing covariates, unidentifiability problems could be solved by abandoning the hazard ratio in favour of the accelerate failure time (AFT) framework, an ordinary regression approach of log(survival time) on covariates with direct physical interpretation, for interpretation of covariate effects in survival analysis with random heterogeneity.[11;18] When the underlying intensity is assumed as Weibull-distributed, the AFT model can give the unchanged regression coefficients on known factors and redistribute the frailty into dispersion.[18]

In multivariate failure time analysis, the idea of the frailty model is to specify independence conditional on a set of unobserved or latent variables, then to complete the model by averaging over an assumed distribution for the latent variables. This can also tackle the problem of unidentifiablity in univariate failure time analysis. Fully parametric versions of the models with parametric specification of the underlying intensity or piecewise constant baseline hazards,[6] and the semiparametric version of the model introducing random effects in

a Cox proportional regression have both been proposed.[7] The latter is widely adopted for its lack of a distribution assumption for the baseline hazard.

The ability to estimate models reliably increases dramatically when multivariate survival times are available, giving more degrees of freedom for parameter estimation.

Large studies based on the Scandinavian Twins Registry data investigated the correlation between twins at different levels, MZ (monozygotic) and DZ (dizygotic), and the relative magnitude of genetic and environmental influences on frailty variables. The applications include the influence of major genes on individual frailty and longevity,[19] familial aggregation of breast cancer in twins,[20] genetic and environmental influences on susceptibility to heart disease,[21] genetic analysis of duration of human logevity,[22-25] the association between cancer incidence rates among the MZ and among the DZ pairs,[26] and the effects of environmental and genetic factors on the mortality from coronary heart disease.[27]

Such models can be used for an unobservable genetic or early environmental effect existing with individuals in sibling groups, or for an environmental effect if individuals are grouped by households. Mack et al[20] studied the familial aggregation of lung cancer in relation to family smoking habits. Siegmund et al[28;29] applied the frailty approach to investigate the effect of latent genetic and environmental risk factors on hazard functions in nuclear families. Klein[30] investigated the risks of smoking and cholesterol levels, adjusting for potential random effects using the data of the Framingham Heart Study. In the same study, Wassell and Moeschberger[31] considered the 'pair-wise' covariate information in the dependence parameter of the bivariate survival function of two events, first detection of hypertension and first cardiovascular disease event, within an individual.

In some applications, measurements were taken in the same subjects but at different anatomical sites. Mache and Chevret[32] developed an approach based on the marginal hazard model and the frailty model to assess the effect of photocoagulation in delaying the onset of blindness, as well as the dependence between the two eyes of blindness times, within patients with diabetic retinopathy. Robertson and Ranstam[33] used a shared gamma frailty model to model the risk of knee prostheses after arthroplasty, taking into account the correlations between bilateral operations in the same patient.

Alternatively, the marginal hazard model can be used for multivariate survival analysis. In this approach, marginal distributions of survival time are modelled as Cox regression, giving population average regression coefficients, but with a correlation structure to be estimated within clusters of related survival time.[34]

1.2.3 Mover-Stayer models

The Mover-stayer model may be considered as a special case of frailty models, where each person has a susceptibility variable (the frailty) which has a Bernoulli distribution. Alternatively it may be thought of as a mixture of two distinct populations, one of which has an unknown hazard of progression (mover), another a hazard rate known to be zero (stayer). The task of estimation is to obtain estimates of the proportion of movers in the population and the parameter pertaining to the non-zero hazard in these movers.

The early development and application is in the discrete-time mover-stayer model, in which movers are frequently assumed to follow a first-order Markov process with unknown transition matrix $\mathbf{M}$, and stayers will remain in their initial state throughout the whole observation period with an identity transition matrix. Let $\mathbf{S}$ denote the diagonal matrix with

ith diagonal element S_i as the fraction of being stayer in state i. Then the jth step transition matrix $\mathbf{P}^{(j)}$ can be expressed as

$$\mathbf{P}^{(j)} = \mathbf{S} + (\mathbf{I} - \mathbf{S})\mathbf{M}^j .$$

Frydman[35] demonstrated maximum likelihood estimation in the discrete-time mover-stayer model directly from the maximum likelihood equations and developed a recursive method of computation. Sampson[36] discussed the approach and that for a discrete time finite-state Markov chain model by using an example of job transitions for unskilled workers. Fuchs and Greenhouse[37] proposed the EM (expectation-maximisation) algorithm as an alternative approach to maximum likelihood estimation in the mover-stayer model. Swensen[38] derived the profile likelihood function to establish a simple necessary and sufficient condition for the maximum likelihood estimator.

Such a simple mover-stayer model in continuous time with a single exponential distribution for the movers and a single Bernoulli random variable with the probability p of being a mover runs into the problem that the likelihood is monotonic increasing in p except in the case of only censored data, and no events. Using a different distribution of time to movement or increasing the complexity of the model with further parameters might solve the problem of monotonic increasing likelihood in p but, depending on the data available, might also give rise to identifiability issues.

Since the first introduction by Blumen, Kogan, and McCarthy,[39] the mover-stayer model has been frequently used to model various forms of dynamics in the social and economical sciences such as occupational mobility,[40-44] residence mobility,[45;46] consumers' brand preferences,[47] income dynamics,[48] analysis of remarriage,[49] and political tendency.[50] In

biomedical studies, the applications of mover-stayer models include investigation of the changes in depression status over time in a mental health survey;[51] the HIV/AIDS epidemic modelling by means of simulation;[52-54] or in terms of the survivor function[55] considering the heterogeneous susceptibility of HIV infection among populations. It has also been used to investigate dedifferentiation and tumour progression in breast cancer via a quasi-likelihood method.[56] Albert treated subjects with dynamic changes between the cut-off point in the scale of defining chronic disease as movers. This demonstrated the importance of accounting for measurement error in estimating prevalence and incidence of a disease which is diagnosed based on dichotomising a continuous marker variable such as blood pressure, fasting blood glucose or respiratory function.[56;57]

1.3 Heterogeneity/ frailty models in cancer screening

Usually, heterogeneity in cancer screening exists on two levels, between screened and unscreened groups, and among screened subjects. The former is a target of cancer screening evaluation methods, while the failure to consider the latter one can distort the quantification of disease natural history that can be used in screening policy making.

Comparison between screened and unscreened groups outside of the randomised trial setting can be difficult because of length bias, lead time bias, and selection or healthy volunteer bias.[58] Length bias and healthy volunteer bias are particular potential applications of frailty models. Researchers have spent much effort investigating such problems and developing methods to make adjustments.[59] Studies by means of computer simulation have been frequently applied in cancer screening evaluation.[60;61]

Heterogeneity among screened subjects is a crucial problem in analysing cancer screening data, especially when the screen aims to detect a premalignant lesion which is more likely to be non-susceptible than other screening targets, such as preclinical invasive cancer. However, formal heterogeneity or frailty models are rarely addressed in this area in the published literature. Nevertheless, screening data hold the advantage of more degrees of freedom, particularly in the design of a randomised screening trial, to reduce the problem of unidentifiability. Researchers have applied a continuous-time mover-stayer mixture of Markov chain models using a quasi-likelihood approach to a randomised mammography screening trial, to investigate whether the population of breast cancer was a mixture of tumours with and without the potential for the deterioration of the malignancy grade.[56;60;62]

In cervical cancer screening, heterogeneity is complicated due to the possibility of regression

of pre-invasive lesion. Traditionally, estimates for the proportion of regression were based on follow-up studies of untreated carcinoma in situ cases.[63] van Oortmarssen and Habbema[64] proposed a five-state (normal, pre-invasive, preclinical invasive, screen-detected and clinical invasive cancer) stochastic model to fit the screening data from the British Columbia cohort study. In their model, time between normal and pre-invasive state was assumed to be Weibull distributed (assumed the same distribution in both progression and regression ways), and for parsimony purpose a number of simplifications were made: parameters from normal and pre-invasive, and to screen-detected state assumed to be piecewise constant with age, and the fixed time between pre-clinical and clinical invasive cancer.

1.4 Purpose of this thesis

Investigation of heterogeneity in the propensity for screen-detected disease to progress is the main target of this project. To investigate the heterogeneity of progression capability of ductal carcinoma in situ (DCIS), we fit data from mammography screening firstly by a simple deterministic approach with various simplifying approximations, and also by a Markov process, in which observed data are a mixture of two latent separate states, progressive and non-progressive DCIS. The model construction and results are described in Chapter 2.

In Chapter 3 a novel method to describe the behaviour of movers or stayers in terms of prognostic factors is applied to a case-cohort study of progression to colorectal cancer of polyps after polypectomy. This uses a two-compartment model. The first compartment has dependent variable the dichotomous potential to progress in the absence of treatment. The second compartment has dependent variable actual progression and time to progression for those which have such a potential.

In addition to picking up invasive carcinoma in the preclinical phase, screening can detect preinvasive and premalignant lesions, such as adenomatous polyps for colorectal cancer,[65;66] and squamous intraepithelial lesions for cervical cancer.[67] Again, heterogeneity of malignant transformation might prevail in the premalignant lesions. In chapter 4, we combine progressive Markov models with constant hazard rates in sequence throughout the process of disease natural history within subjects, with a frailty model for continuous heterogeneity and with a mover-stayer model for dichotomous heterogeneity for progression between sequential homogeneous models. We use this method to estimate the parameters of progressive disease natural history for colorectal cancer.

We introduce latent variables in the multi-state disease process in chapter 5 to tackle the situation of unknown or unobserved information, sometimes due to a lack of knowledge of the importance of a particular variable when data are being collected, or sometimes the information is only meaningful or observable for some selective subjects. The example used is calcification type in mammography, which was only known for tumours of size 1-14 mm, in mammographic screening.

In Chapter 6, a variety of disease progressive models were fitted to investigate the heterogeneity of progression of abnormalities detected in cervical cancer screening. A basic model taking length bias into account suggests that the low estimated malignant transition rates are due to ignoring heterogeneity of susceptibility. A further mover-stayer model is used to deal with the heterogeneity among premalignancies, also incorporating measurement error in the model. To assess cervical cancer screening, a computer simulation technique is applied to compare screening with different regimes. This is used to estimate benefits and costs of different screening frequencies in terms of smears required per invasive cancer prevented.

In Chapter 7, we summarise the particular clinical conclusions, and the generic methodological conclusions. Suggestions for further work, both clinical and methodological, are made.

Chapter 2 Heterogeneity of Progression Capability of Ductal Carcinoma In Situ (DCIS)

2.1 Introduction

Since the inception of mammographic screening programmes for breast cancer, concerns have been expressed about the possibility of overdiagnosis of breast cancer, in particular of ductal carcinoma in situ (DCIS),[68;69] a noninvasive or preinvasive lesion in the breast. A dramatic increase in incidence of DCIS has been observed since then. According to the Surveillance, Epidemiology and End Results Registry, the increase of age-adjusted incidence rate of DCIS among women was 314% between 1983 and 1993 in the United States. For invasive breast cancer, the increase was only 16%.[69] The increasing trend was contemporaneous with the increase in mammographic screening since the early 1980s. High percentages of DCIS among screen-detected tumours have been reported, as has an increased incidence of DCIS since the advent of screening. The proportion of DCIS of all newly diagnosed breast cancer increased from 3% between 1973 and 1980 to 15.5% in 1996, and is estimated as nearly 20% in 2000.[70] The percentage of DCIS among screen-detected breast cancer decreases with age from 28% for women aged 40-49 years to 16% for women aged 70-84 years.[71] The percentage in prevalence mammographic screening (i.e. the first screen of a previously unscreened population) is even higher and also declines with age from 43% for women aged 40-49 to 19% for those aged 60-69.[69]

Various interpretations of these results have appeared in the literature ranging from the opinion that this is mainly overdiagnosis and likely to cause more harm than benefit, to the position that detection of DCIS is the ideal and that a high rate of DCIS represents a large number of invasive cancers avoided. Thus, the critical question is: 'Is DCIS an obligate precursor lesion of invasive breast cancer?'.

Researchers have given evidence of similarities of risk factors and of genetic markers in DCIS and invasive breast cancer. Kerlikowske et al[72] found that family history of breast cancer and nulliparity or late age of delivering of first child significantly increased the risk of both mammographically detected DCIS and invasive breast cancer among women aged older than 50 years. For younger women an elevated BMI was also associated with decreased risk of DCIS[72;73] and the same trend bordering on statistical significance was observed for invasive breast cancer.[72] Biological evidence showed the progression of genetic abnormalities, including change in estrogen receptor (ER) levels, expression of oncogene c-erb B-2, tumour suppressor gene p53 and loss of heterozygosity at multiple chromosomal loci, from atypical ductal hyperplasia to low-grade DCIS, to high-grade DCIS, and finally to invasive ductal carcinoma.[70] The evidence also comes from the similar distribution of location in the breast of DCIS and of invasive cancer.[68;72] The earlier age of diagnosis and smaller tumour size of DCIS compared with those of invasive breast cancer also show a potential precursor role of DCIS for invasive breast cancer.[72]

By contrast, some argue that the majority of DCIS cases do not necessarily progress to clinically significant invasive breast cancer and therefore that DCIS cannot be considered as a precursor lesion. The autopsy studies have typically observed a high prevalence of DCIS, even in women who were not diagnosed with breast cancer during their lifetime. Welch and Black[74] observed a prevalence of 9%. Nielsen et al[75;76] found a prevalence of 14% from series of consecutive autopsies of women aged over 20 years.

From the results of follow-up studies for women diagnosed with benign breast disease by misclassification and left untreated, but which were actually DCIS cases, Betstill et al[77] found 6 invasive cancers (60%) occurred out of 10 DCIS cases after an average of 9.7 years follow

up, Rosen et al[78] updated the data and found 9 invasive cancer cases (60%) out of 15 DCIS cases after 21.6 year follow up. Page et al[79;80] found 9 ipsilateral invasive breast cancer (32%) out of 28 misclassified cases taking biopsy only after 30 year follow up, and Eusebi et al[81] reported 9 ipsilateral invasive breast cancer (11%) from 80 cases after 17.5 year follow up.

However, the above results involve uncertainties about the representative nature of the varieties of DCIS diagnosed by varying detection methods. First, the results of autopsy studies do not necessarily reflect the DCIS in the living population or mammographically detectable DCIS in vivo. If around 9% of women at autopsy are found to have DCIS,[75] but detection rates at screening are less than 1 per thousand,[71] it is likely that screen-detectable DCIS in vivo is not the same clinical entity as autopsy diagnosed DCIS. In addition, Evans et al[82] found a higher proportion of high grade cases among screen-detected DCIS compared to symptomatic cases (69% vs 61%, p=0.08) and a significantly higher proportion of necrotic lesions (87% vs 76%, p=0.008). Both features (high grade and necrosis) confer an increased risk of progressive to invasion. High grade DCIS more often shows abnormal mammographic features than low grade DCIS, which is often mammographically occult. The mammographic calcification found in high grade DCIS and DCIS with necrosis is more characteristic of malignancy. By contrast, the granular-punctate calcifications found in low grade DCIS are non-specific and are more likely to be confused with benign processes. These cases are likely to be the subjects in the follow-up studies.

Percentages of DCIS among screen-detected cases are often quoted in evaluation of screening programmes as an indicator of benefit or harm depending on opinion. It is not clear what the prevalence of DCIS should be, or to what extent variability in percentages of DCIS between screening programmes is due to variation in the absolute rate of DCIS per person screened or to variation in the absolute rate of invasive carcinoma, which also affects the percentage. For

example, a prevalence screen which results in 2 DCIS cases per thousand and 4 invasive carcinoma cases per thousand has 33% DICS, whereas a prevalence screen resulting in 2 cases of DCIS per thousand and 7 invasive cases per thousand will have 22% DCIS. The former is not diagnosing DCIS any more than the latter; it is diagnosing fewer invasive cancers.

Duffy et al[83] found that detection of DCIS in the Swedish Two-County Trial accounted for 5-12% of the deaths prevented, whereas shifting from invasive stage II or worse to invasive stage I accounted for around 65% of the deaths prevented. The contribution of saving lives from breast cancer from detection of DCIS is modest. In addition, the variation between programmes in percentages of cases of DCIS was as much due to invasive carcinoma detection rates as to DCIS detection rates (see Table 2.1). The question remains, however, as to what rates of detection of DCIS should ideally be observed?

Table 2.1 Detection rates per thousand screened of DCIS and invasive cancers at incidence screen and prevalence screen, in the Swedish Two-county Trial, and in more recent service screening programmes

Programme and age range	Prevalence Screen			Incidence screen (where available)		
	DCIS /1,000	Invasive /1,000	% DCIS	DCIS /1,000	Invasive /1,000	% DCIS
Two-County, 50-69	0.7	6.0	11	0.5	3.3	13
UK, 50-64	1.1	4.9	18	0.6	3.2	16
Netherlands, 50-69	0.9	5.5	14	0.5	2.9	15
Belgium, 40-69	1.0	7.1	12	-	-	-
South Australia, 40+	1.3	5.8	18	0.6	2.8	18
New York, 98% 35-69	0.8	4.4	15	0.3	1.5	20
San Francisco, 50-59	1.8	4.2	30	-	-	-
San Francisco, 60-69	2.2	9.8	18	-	-	-

It is generally agreed that a proportion of DCIS would not have progressed to invasive carcinoma in the absence of screening. This proportion, however, is unknown, and cannot be observed due to the interruption of natural history by excision. Also, as stated above, long term results in DCIS previously untreated for benign condition are qualified by the non-representative nature of such cases. Therefore, the only way to estimate the proportion of progressive DCIS cases is by statistical modelling based on the numbers of DCIS and invasive cases detected at screening and on the numbers of breast cancers arising clinically between screenings.

2.2 Objectives

In the present chapter we use data from the Swedish Two-County Trial and from service screening programmes to:

(1) derive tentative estimates of what DCIS detection rates should be typically observed;

(2) describe the typical range of absolute detection rates of DCIS in mammographic screening programmes;

(3) estimate the proportion of DCIS detected at prevalence and incidence screens which is "overdiagnosis", ie which would not have progressed to invasive disease if left untreated.

We begin with a simple deterministic model incorporating various constraints. Subsequently, we fit a six-state stochastic model estimating five parameters of incidence and progression simultaneously.

2.3 Material and methods

2.3.1 Data

We use data from the Swedish Two-County trial, a randomised controlled trial with 77,080 women aged 40-74 randomly assigned to screening invitation and resulting in 1,426 breast cancers during the trial, and 55,985 women assigned to no invitation and resulting in 1,042 breast cancers.[84] In the present study, only the data on DCIS and invasive breast carcinomas detected at prevalence and the first subsequent incidence screen, interval cancers detected between prevalence and the first subsequent incidence screen and carcinoma-free cases for women aged 40-69 were used. In addition, the corresponding figures from service screening programmes in the UK,[85] Netherlands,[86] South Australia,[87] and New York[88] were extracted from published papers (Table 2.2). Interval cancer data for these programmes were not available.

Table 2.2 Numbers of DCIS, invasive carcinoma and women carcinoma-free at prevalence and the first incidence screen in the Swedish Two-County Trial, UK, Netherlands, South Australia and UK programmes

	Detection mode	Prevalence screen	First incidence screen
Swedish Two-County	DCIS	8	7
40-49	Invasive carcinoma	31	39
	Carcinoma-free	18456	16396
	Interval cancer*	20	
Swedish Two-County	DCIS	15	12
50-59	Invasive carcinoma	87	49
	Carcinoma-free	21457	18731
	Interval cancer*	25	
Swedish Two-County	DCIS	17	8
60-69	Invasive carcinoma	167	81
	Carcinoma-free	20395	16372
	Interval cancer*	38	
UK	DCIS	2767	173
	Invasive carcinoma	12323	925
	Carcinoma-free	2520526	227503
Netherlands	DCIS	908	383
	Invasive carcinoma	5548	2223
	Carcinoma-free	1077844	791790
	Interval cancer*	760	
South Australia	DCIS	94	12
	Invasive carcinoma	439	61
	Carcinoma-free	75573	21433
New York	DCIS	42	17
	Invasive carcinoma	230	67
	Carcinoma-free	52378	45839

* Numbers of interval cancers in Swedish Two-County trial and the Netherlands programme are adjusted by the compliance rate to represent the subgroup of those attending the first subsequent screen

2.3.2 Statistical models and methods

We analysed the data in two ways. Firstly, we used a simple deterministic approach, using various simplifying approximations, and secondly, by explicitly fitting a Markov process to occurrence and progression of in situ and invasive carcinoma.

2.3.2.1 Deterministic model

Suppose that two types of DCIS, progressive and non-progressive, follow different natural history models (A) and (B), respectively (Figure 2.1). In model (A), there is progression from normal to DCIS, DCIS to preclinical invasive disease, and preclinical invasive disease to clinical invasive disease. In model (B), the DCIS does not progress further and may indeed regress. Observed DCIS cases at screening are a mixture of these two types, but the proportion of progressive cases and the status of any individual case are unknown.

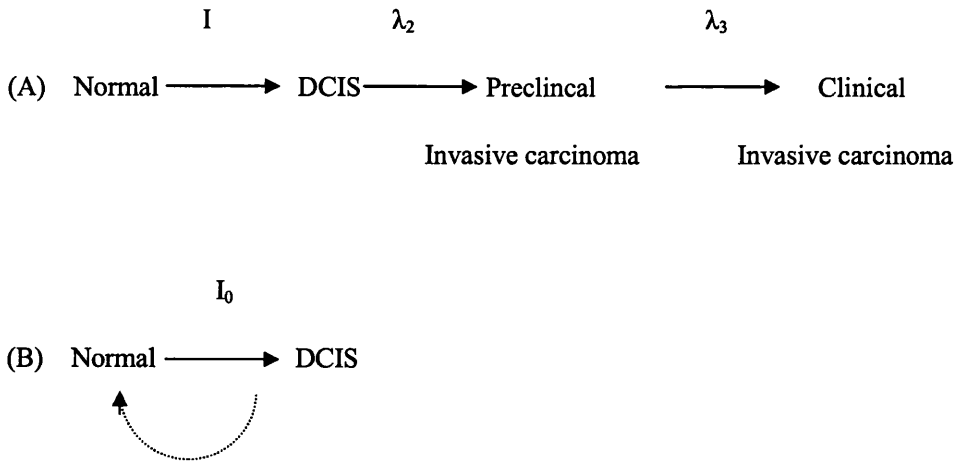


Figure 2.1 Natural history models involving progressive DCIS (A) and non-progressive DCIS (B).

In addition, we define a to be the mean number of years since non-progressive DCIS was born, I_0 to be the annual incidence of non-progressive DCIS, S_1 to be the screening sensitivity to

DCIS, and S_2 to be sensitivity to invasive carcinoma.

Before estimation, some simplifying assumptions are made as follows:

(1) The incidence of preclinical progressive DCIS, preclinical invasive carcinoma, and clinical invasive carcinoma in the absence of screening are equal. This represents a steady state situation in which all clinical invasive cancer must pass through the preclinical in situ and invasive state. They are approximated from incidence of invasive carcinoma in the control group before the first screen in the Swedish Two-County trial. We assume uniform annual incidence.

(2) The annual progression rate from preclinical invasive carcinoma to clinical disease follows an exponential distribution and the progression rate and the sensitivity to preclinical invasive cancer are known from previous work. Progression from in situ to invasive carcinoma follows an exponential distribution with an unknown rate to be estimated from the data.

(3) a is very large compared with screening interval t.

(4) There is no overdiagnosis of invasive carcinoma.

(5) Sensitivity to progressive and non-progressive DCIS is equal.

In this analysis, we use only screen-detected cases, not interval cancers.

At prevalence screen, the probability of detecting DCIS is

$$P_1(DCIS) = \left(\frac{I}{\lambda_2} + I_0 a\right) S_1 = p_1 \qquad [2.1]$$

That is, the sum of incidence of progressive and non-progressive DCIS multiplied by their respective sojourn times, and by screening sensitivity.

At prevalence screen, the probability of detecting preclinical invasive carcinoma is

$$P_1(INV) = \frac{I}{\lambda_3} S_2 = p_2 \qquad [2.2]$$

i.e. the incidence of preclinical invasive carcinoma times its sojourn time, times the screening sensitivity.

At first incidence screen, the probability of detecting DCIS is

$$P(DCIS) = \left(\frac{I\left(1 - e^{-\lambda_2 t}\right)}{\lambda_2} + I_0 t \right) S_1 + I_0 a (1 - S_1) S_1 + \frac{I}{\lambda_2} e^{-\lambda_2 t} (1 - S_1) S_1 = p_3 \qquad [2.3]$$

The first component represents newly arising progressive and non-progressive DCIS. The two latter components represent non-progressive and progressive cases missed at first screen. The probability of detecting preclinical invasive carcinoma at first incidence screen can be represented as follows:

$$P(INV) = \frac{I}{\lambda_3} e^{-\lambda_3 t} (1 - S_2) S_2 + \frac{I\left(1 - e^{-\lambda_3 t}\right)}{\lambda_3} S_2 + \frac{I S_2 (1 - S_1)}{\lambda_2} \left\{ \int_0^t \lambda_2 e^{-\lambda_2 v} e^{-\lambda_3 (t - v)} dv \right\}$$

The first component represents invasive cases missed at prevalence screen, the second newly arising invasive cases and the third progressive DCIS missed at first screen and progressing to invasive during the interval. Our simplifying "steady state" assumption of equal incidence of preclinical invasive and progressive in situ cases obviate the need to model incidence of the later and progression to the former in the second component. This simplifies to

$$= IS_2\left\{\frac{1-S_2e^{-\lambda_3 t}}{\lambda_3}+\frac{(1-S_1)\left(e^{-\lambda_2 t}-e^{-\lambda_3 t}\right)}{\lambda_3-\lambda_2}\right\} \qquad [2.4]$$

$$= p_4$$

Suppose N_1 women attend the prevalence screen, and n_{11} and n_{12} are detected as DCIS and invasive carcinoma, respectively. The numbers of attending women, screen-detected DCIS and invasive carcinoma cases at the first subsequent screen are N_2, n_{21}, and n_{22}. Based on the probability formulae derived above, the total likelihood is:

$$L = p_1^{n_{11}} p_2^{n_{12}} (1-p_1-p_2)^{N_1-n_{11}-n_{12}} p_3^{n_{21}} p_4^{n_{22}} (1-p_3-p_4)^{N_2-n_{21}-n_{22}} \qquad [2.5]$$

Based on the assumptions, I, λ_3, and S_2 are regarded as constants. This means that there are four parameters,I_0, a, λ_2, and S_1, left in the likelihood function to be estimated. The maximum likelihood estimates for I_0, a, λ_2, and S_1 were obtained using Newton-Raphson optimization. SAS/IML 6.12 software was used.

2.3.2.2 Markov models

We consider the six-state continuous-time Markov model depicted in Figure 2.2. In this model, progressive and non-progressive DCIS are taken into account simultaneously. Note that state (5), no tumour apparent after non-progressive DCIS regression is not treated as a return to normal, but as a separate absorbing state. This is done partly to assist in estimation but also to reflect the fact that the natural history of non-progressive DCIS is not yet established. For example, those in whom a non-progressive DCIS has regressed might be at increased risk of a new primary tumour thereafter. Note that in this model, DCIS, whether progressive or not, denotes ductal carcinoma in situ which has given rise to calcifications and is therefore screen-detectable. It does not include undetectable DCIS.

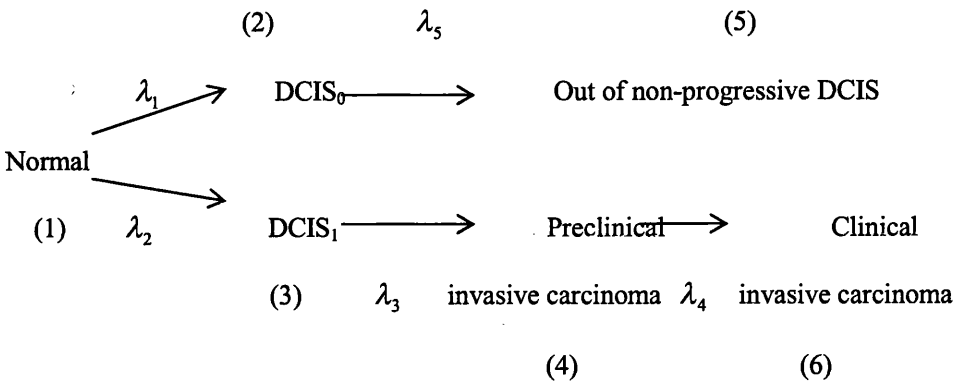


Figure 2.2 Natural history of breast cancer including progressive and non-progressive DCIS

The transition rates of the six-state model above can be expressed as an intensity matrix:

$$Q = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{pmatrix} -(\lambda_1+\lambda_2) & \lambda_1 & \lambda_2 & 0 & 0 & 0 \\ 0 & -\lambda_5 & 0 & 0 & \lambda_5 & 0 \\ 0 & 0 & -\lambda_3 & \lambda_3 & 0 & 0 \\ 0 & 0 & 0 & -\lambda_4 & 0 & \lambda_4 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \qquad [2.6]$$

λ_1 and λ_2 represent the DCIS incidence rates for non-progressive and progressive DCIS, respectively. λ_3 and λ_4 are the transition rates from progressive DCIS to the invasive preclinical phase, and from preclinical invasive to clinical disease, respectively. λ_5 is the annual transition rate from non-progressive DCIS to no apparent tumour.

Given the transition intensity matrix in [2.6], transition probabilities from state i to state j in a time interval, x, can be expressed as $P_{ij}(x)$. The derivation of transition probabilities is based on the solution of the Kolmogorov equations[89;90] or by integration.

From the Kolmogorov equations,

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{Q}, \qquad [2.7]$$

which has the unique solution $\mathbf{P}(t) = \exp(\mathbf{Q}t)$ subject to the boundary condition $\mathbf{P}(0) = \mathbf{I}$. Using canonical decomposition, for the given set of parameters, λ_1 to λ_5 in the present study, $\mathbf{Q}$ has distinct eigenvalues $d_1, \ldots, d_6$ and $\mathbf{A}$ is the k×k matrix whose jth column is a right eigenvector corresponding to d_j, then $\mathbf{Q} = \mathbf{ADA}^{-1}$, where $\mathbf{D} = diag(d_1,...,d_k)$. Then

$$\mathbf{P}(t) = \mathbf{A}diag\left(e^{d_1 t},.....,e^{d_6 t}\right)\mathbf{A}^{-1} \qquad [2.8]$$

Using integration way, for example, the probability of having preclinical invasive cancer at a prevalence screen at a given age is

$$\begin{aligned} P_{14}(Age) &= \int_0^{age} \lambda_2 e^{-\lambda_2 x} e^{-\lambda_1 x} \int_0^{age-x} \lambda_3 e^{-\lambda_3 y} e^{-\lambda_4 (age-x-y)} dy dx \\ &= \frac{\lambda_2 \lambda_3}{\lambda_4 - \lambda_3}\left\{\frac{e^{-(\lambda_1+\lambda_2)age} - e^{-\lambda_3 age}}{\lambda_3 - \lambda_1 - \lambda_2} - \frac{e^{-(\lambda_1+\lambda_2)age} - e^{-\lambda_4 age}}{\lambda_4 - \lambda_1 - \lambda_2}\right\} \end{aligned} \qquad [2.9]$$

In the present study, we constructed the likelihood from expression [2.8], the solution of the Kolmogorov equations, by numerical calculation when the iterations of optimisation were performed. In this case, a closed form of progression probability in a given time t is not necessary. This method is especially useful when a transition matrix involves more than two regressive modes and a closed form expression does not exist.

The probabilities used for maximum likelihood estimation in this six-state model are shown in Table 2.3. Note that in this more formal representation, it is necessary to take account of the fact that those with a history of clinical breast cancer prior to the start of screening are

excluded. This is the reason for the denominators in the first screen probabilities in Table 2.3. The probabilities are conditional on being either normal, with DCIS, or with preclinical invasive disease at first screen. In this analysis, we use interval cancer data when available.

Table 2.3 Transition probabilities by detection mode at prevalence and the first subsequent screen in six-state Markov model

Detection Mode	Transition Probability for models with interval cancer	Transition Probability for models without interval cancer
Prevalence Screen		
DCIS	$\frac{P_{12}(Age)+P_{13}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}$	$\frac{P_{12}(Age)+P_{13}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}$
Invasive carcinoma	$\frac{P_{14}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}$	$\frac{P_{14}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}$
Carcinoma-free	$\frac{P_{11}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}$	$\frac{P_{11}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}$
Subsequent screen		
DCIS	$P_{12}(x)+P_{13}(x)$	$[P_{12}(x)+P_{13}(x)]\times Coverage$
Invasive carcinoma	$P_{14}(x)$	$P_{14}(x)\times Coverage$
Carcinoma-free	$P_{11}(x)$	$P_{11}(x)\times Coverage$
Interval Cancer	$P_{16}(x)$	---
Non-attender	---	$[P_{11}(x)+P_{12}(x)+P_{13}(x)+P_{14}(x)]$ $\times(1-Coverage)+[P_{15}(x)+P_{16}(x)]$

N_1, n_{11}, n_{12}, N_2, n_{21} and n_{22} stand for the same meaning in the deterministic model. The number of interval cancers between prevalence screen and the first subsequent screen is n_{ic}. Based on the probability formulae in table 2.3, the total likelihood, including data on interval cancers is:

$$L=\left(\frac{P_{12}(Age)+P_{13}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}\right)^{n_{11}}\times$$
$$\left(\frac{P_{14}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}\right)^{n_{12}}\times \qquad [2.10]$$
$$\left(\frac{P_{11}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}\right)^{N_1-n_{11}-n_{12}}\times$$
$$(P_{12}(x)+P_{13}(x))^{n_{21}}\times(P_{14}(x))^{n_{22}}\times(P_{11}(x))^{N_2-n_{21}-n_{22}}\times(P_{16}(x))^{n_{ic}}$$

where age is the age of women attending the prevalence screen, and x is the interval between the prevalence screen and the first subsequent screen. In the service screening programmes, there were no data on interval cancers available. In this case, we use a complement probability including total likelihood from women not attending the first subsequent screen (including women not yet invited thereto) and all interval cancers appearing between the prevalence screen and the first subsequent screen. Coverage is $N_2/(N_1-n_{11}-n_{12})$, the proportion of those attending the first screen who also attend the second. Then, the total likelihood for screening programmes without interval cancers is:

$$L=\left(\frac{P_{12}(Age)+P_{13}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}\right)^{n_{11}}\times$$
$$\left(\frac{P_{14}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}\right)^{n_{12}}\times$$
$$\left(\frac{P_{11}(Age)}{P_{11}(Age)+P_{12}(Age)+P_{13}(Age)+P_{14}(Age)}\right)^{N_1-n_{11}-n_{12}}\times$$
$$[(P_{12}(x)+P_{13}(x))\times Coverage]^{n_{21}}\times[P_{14}(x)\times Coverage]^{n_{22}}\times[P_{11}(x)\times Coverage]^{N_2-n_{21}-n_{22}}\times$$
$$\{[P_{11}(x)+P_{12}(x)+P_{13}(x)+P_{14}(x)]\times(1-Coverage)+[P_{15}(x)+P_{16}(x)]\}^{N_1-n_{11}-n_{12}-N_2}$$

[2.11]

The maximum likelihood estimates were obtained by using Newton-Raphson optimisation. SAS/IML 6.12 software was used. 95% confidence intervals were calculated using variance estimated from the inverse Hessian matrix.

Note that the coverage at second screening is not to be confused with compliance. At the time of the reports in several of the service screening programmes, the first screening was completed but the second still under way. Thus, those who had not attended the second screening were a mixture of those choosing not to attend and those who had not yet been invited to a second screening.

2.4 Results

2.4.1 Deterministic models

We were unable to obtain convergence from the Swedish Two-County trial data for ages 40-49, probably because of the relatively small number of cancers. The constrained values for incidence of progressive DCIS, rate of progression of invasive preclinical disease to clinical disease and sensitivity to invasive disease were 0.0021, 0.2702, and 1.00 for women aged 50-59, 0.0028, 0.2381, and 1.00 for women aged 60-69, and 0.0024, 0.2542, and 1.00 for women aged 50-69, from Tabar et al.[84]

The estimates from the Swedish Two-County trial for 10-year age groups were unstable results, and are therefore not reported. Instead, we estimated the parameters for the Two-County Trail for the age group 50-69 as a whole.

Table 2.4 shows the estimated annual incidence rates of non-progressive DCIS, sojourn time of non-progressive DCIS, annual transition rates from progressive DCIS to preclinical invasive carcinoma, and sensitivity to DCIS from the various sources. Estimates of the sojourn time ranging from 3 to 7 years of non-progressive DCIS were observed in the Swedish trial, UK, Netherlands, and South Australia. However, a longer sojourn time of non-progressive DCIS, 13.17 years, was estimated in the New York programme. The estimated annual non-progressive DCIS incidence rates varied between 1.00 and 1.67 per 10,000 per year, except that of New York programme, 3.80 per 10,000 per year. The estimates of annual transition rates from progressive DCIS to preclinical invasive carcinoma ranged from 8.38 (average time to progression 1.4 months) in the New York programme, to 10.94 (average time to progression 1.1 months) in the Netherlands programme, and results generally indicated a short sojourn time of progressive DCIS. Sensitivities to DCIS were consistently

estimated as 1. Pooled estimates which are weighted by the number of breast carcinoma cases for annual incidence and mean sojourn time of non-progressive DCIS, annual transition rates from progressive DCIS to preclinical invasive carcinoma, and the sensitivity to DCIS were 1.51 per 10,000 per year, 5.04 years, 9.78 and 1, respectively. However, estimated sojourn time of non-progressive DCIS was unstable and a confidence interval could not be estimated in many cases. We therefore attempted an alternative estimation algorithm.

Table 2.4 Estimated parameters based on the deterministic model

	Parameter	Estimate	95% CI**
Swedish Two-County	I_0	1.10×10^{-4}	0~0.0420
50-69	a	4.48	0.0041~4841.80
	λ_2	9.30	0.0300~2886.88
	S_1	1.00	0~1.00
UK	I_0	1.67×10^{-4}	N/E
	a	4.98	N/E
	λ_2	9.20	N/E
	S_1	1.00	N/E
Netherlands	I_0	1.28×10^{-4}	0.0001~0.0002
	a	4.74	3.89~5.80
	λ_2	10.94	9.59~12.47
	S_1	1.00	N/E
South Australia	I_0	1.41×10^{-4}	2.13×10^{-5}~9.39×10^{-4}
	a	6.76	N/E
	λ_2	8.94	1.23~64.92
	S_1	1.00	N/E
New York	I_0	3.80×10^{-4}	N/E
	a	13.17	N/E
	λ_2	8.38	N/E
	S_1	1.00	N/E
Overall*	I_0	1.51×10^{-4}	N/E
	a	5.04	N/E
	λ_2	9.78	N/E
	S_1	1.00	N/E

* Weighted by number of breast carcinoma cases

** Calculated by Delta method for I_0, a, andλ_2, and truncated method for S_1

Table 2.5 shows the results of profile likelihood estimation of a, with Newton-Raphson optimized estimates of the remaining parameters constrained on a. Compared with the results shown in table 2.4, the estimates of annual incidence rates of non-progressive DCIS became much smaller. However, the estimates of sojourn time of non-progressive DCIS were greater than those in table 2.4. Slower transitions from progressive DCIS to preclinical invasive carcinoma were estimated. The estimates of annual transition rates from progressive DCIS to preclinical invasive carcinoma ranged from 5.09 (average time to progression 2.4 months) in the Swedish Two-County trial, women aged 50-69, to 8.05 (average time to progression 1.5 month) in the Netherlands programme. Again, 100% sensitivities were consistently estimated in this approach. There were still problems in interval estimation.

Table 2.5 Estimated parameters based on the deterministic model using the profile likelihood in estimation for sojourn time of non-progressive DCIS

	Parameter	Estimate	95% CI**
Swedish Two-County	a	9	N/E[&]
50-69	I_0	3.04×10^{-5}	5.50×10^{-7}~1.68×10^{-3}
	λ_2	5.09	2.56~10.14
	S_1	1.00	0~1
UK	a	6	5~7
	I_0	1.10×10^{-4}	7.14×10^{-5}~1.70×10^{-4}
	λ_2	5.52	4.08~7.49
	S_1	1.00	0.76~1.00
Netherlands	a	6	3~8
	I_0	8.80×10^{-5}	N/E
	λ_2	8.05	N/E
	S_1	1.00	N/E
South Australia	a	12	N/E
	I_0	6.73×10^{-5}	8.42×10^{-6}~5.38×10^{-4}
	λ_2	5.82	1.65~20.50
	S_1	1.00	0~1
New York	a	23	2~30
	I_0	2.02×10^{-5}	0.0000~6.1619
	λ_2	7.48	0.12~457.38
	S_1	1.00	0~1
Overall*	a	6.41	N/E
	I_0	9.39×10^{-5}	N/E
	λ_2	6.41	N/E
	S_1	1.00	N/E

[&] Likelihood almost flat giving a 95% CI so wide as to be meaningless

* Weighted by number of breast carcinoma cases

** Calculated by Delta method for I_0, andλ_2, and truncated method for S_1

2.4.2 Markov models

Table 2.6 shows results from estimation based on the six-state Markov model. Note that the exponential distribution of time to transition implies that the inverse of the estimated transition rate from a state is the estimated mean time spent in the state (mean sojourn time). Thus, for example, for the Swedish Two-County trial, women aged 50-59, the estimated sojourn time of invasive preclinical cancer before transition to clinical symptomatic disease is 1/0.40=2.50 years, and the mean sojourn time of non-progressive DCIS is estimated as 1/0.0617=16.21 years.

The annual incidence rates for non-progressive DCIS ranged from 7.22 per million per year in the Swedish Two-County trial, women aged 40-49, to 7.27 per 100,000 per year in the UK programme. The annual incidence rates for progressive DCIS ranged from 1 to 2.7 per thousand. The shortest estimated average time of progression of DCIS to invasive disease was 2 months in the Swedish Two-County trial for women aged 60-69, and the longest was 5.22 months in the New York programme. The estimated average sojourn time of non-progressive DCIS ranged from 6 years in the UK programme to 40 years in the Netherlands programme. Pooled estimates derived from weighted averages of the annual incidence rates for non-progressive DCIS and progressive DCIS, the sojourn time of non-progressive and progressive DCIS and sojourn time of preclinical invasive carcinoma were 1.11 per 100,000, 2.1 per thousand, 30 years, 3 months, and 2.5 years, respectively. Variance estimates on parameters pertaining to non-progressive DCIS could not be obtained, partly due to the very low incidence and therefore to the lack of data on which to base estimation. The profile likelihood was therefore used to calculate 95% confidence intervals on parameters pertaining to non-progressive DCIS. This strategy also failed for the rate of transition from non-progressive DCIS to no apparent tumour in the Swedish Two-County for women aged 40-49 and aged 50-59 because of an almost flat likelihood in this dimension.

Table 2.6 Estimated results of the six-state Markov model

	Parameter	Estimate	95% CI
Swedish Two-County 40-49	Normal → $DCIS_0$	7.22×10^{-6}	$0 - 3.89\times10^{-5}$
	Normal → $DCIS_1$	0.0017	0.0013 – 0.0021
	$DCIS_1$ → Inv	4.93	1.24 – 8.61
	Inv → Clinical	0.80	0.57 – 1.04
	$DCIS_0$ → Out of DCIS	0.0857	N/E
Swedish Two-County 50-59	Normal → $DCIS_0$	9.86×10^{-6}	$0 - 3.54\times10^{-5}$
	Normal → $DCIS_1$	0.0016	0.0013 – 0.0019
	$DCIS_1$ → Inv	2.99	1.22 – 4.76
	Inv → Clinical	0.40	0.30 – 0.49
	$DCIS_0$ → Out of DCIS	0.0617	N/E
Swedish Two-County 60-69	Normal → $DCIS_0$	1.18×10^{-5}	$1.5\times10^{-6} - 2.60\times10^{-5}$
	Normal → $DCIS_1$	0.0027	0.0023 – 0.0032
	$DCIS_1$ → Inv	6.13	1.92 –10.33
	Inv → Clinical	0.33	0.27 – 0.40
	$DCIS_0$ → Out of DCIS	0.0273	0 – 0.2080
UK	Normal → $DCIS_0$	7.27×10^{-5}	$6.57\times10^{-5} - 7.93\times10^{-5}$
	Normal → $DCIS_1$	0.0026	0.0023 – 0.0028
	$DCIS_1$ → Inv	3.87	3.20 – 4.53
	Inv → Clinical	0.52	0.46 – 0.58
	$DCIS_0$ → Out of DCIS	0.1693	0.1546 – 0.1859
Netherlands	Normal → $DCIS_0$	1.07×10^{-5}	$1.06\times10^{-5} - 1.09\times10^{-5}$
	Normal → $DCIS_1$	0.0021	0.00199 – 0.00212
	$DCIS_1$ → Inv	4.19	3.83 – 4.54
	Inv → Clinical	0.39	0.38 – 0.41
	$DCIS_0$ → Out of DCIS	0.0253	0.0235 – 0.0254
South Australia	Normal → $DCIS_0$	2.89×10^{-5}	$1.68\times10^{-5} - 4.27\times10^{-5}$
	Normal → $DCIS_1$	0.0021	0.0015 – 0.0029
	$DCIS_1$ → Inv	3.12	2.08 – 6.22
	Inv → Clinical	0.37	0.25 – 0.52
	$DCIS_0$ → Out of DCIS	0.0511	0.0304 – 0.0892
New York	Normal → $DCIS_0$	1.41×10^{-5}	$5.1\times10^{-6} - 2.54\times10^{-5}$
	Normal → $DCIS_1$	0.0010	0.0008 – 0.0013
	$DCIS_1$ → Inv	2.30	1.44 – 3.17
	Inv → Clinical	0.24	0.17 – 0.31
	$DCIS_0$ → Out of DCIS	0.0388	0.0119 - 0.1092
Overall[&]	* Normal → $DCIS_0$	1.11×10^{-5}	$1.098\times10^{-5} - 1.129\times10^{-5}$
	* Normal → $DCIS_1$	0.0021	0.00202-0.00213
	* $DCIS_1$ → Inv	4.0030	3.7242-4.3025
	* Inv → Clinical	0.3990	0.3857-0.4127
	* $DCIS_0$ → Out of DCIS	0.0332	0.0321-0.0344

^ Likelihood almost flat and a 95% CI so wide as to be meaningless

[&] Weighted average from Swedish Two-County, women aged 50-69, and screening programs of UK, Netherlands, South Australia and New York

* Weighted by the precision of parameters

Table 2.7 shows the probability of non-progressive DCIS, progressive DCIS, preclinical invasive carcinoma and the proportion of non-progressive DCIS at prevalence and at the first subsequent screen derived from the estimates in the six-state Markov model. The larger the proportion of non-progressive DCIS, the more serious the problem of over-diagnosis. Overall, about 20% to 50% of DCIS cases (average 37%) at prevalence screen were estimated as non-progressive. At first incidence screen, the corresponding figure was 3-7%, except for 21% in the UK programme, with an average of 4%.

Table 2.7 Estimated rate per 100,000 of $DCIS_0$, $DCIS_1$, and preclinical invasive carcinoma with the proportion of $DCIS_0$ at prevalence and the first subsequent screen under six-state Markov model

	Prevalence Screen				First subsequent screen			
	$DCIS_0$	$DCIS_1$	Preclinical INV	$DCIS_0$/ DCIS	$DCIS_0$	$DCIS_1$	Preclinical INV	$DCIS_0$/ DCIS
Swedish Two-County								
40-49	8	35	216	19%	1	35	164	3%
50-59	16	54	411	23%	2	54	251	4%
60-69	38	44	821	46%	3	44	473	6%
UK	43	66	490	39%	17	65	369	21%
Netherlands	34	49	520	41%	2	49	259	4%
South Australia	56	67	578	46%	5	67	263	7%
New York	33	45	439	42%	3	45	135	6%
Overall	**30**	**52**	**520**	**37%**	**2**	**52**	**279**	**4%**

Goodness-of-fit tests (Table 2.8) for Markov models in the Swedish Two-County trial, for women aged 50-59 and 60-69 at randomisation, show very good model fitting. The Pearson Chi-squareds on 2 degrees of freedom (7 counts, 5 parameters) were 0.33 and 0.17 for models

with women aged 50-59 and with women aged 60-69, respectively. However, the model seems to fit poorly in the Swedish Two-County trial for women aged 40-49, and the Netherlands programme with Pearson Chi-squared of 8.12 (p=0.017), and 19.20 (p=0.00007), respectively.

Table 2.8 Model fitting and goodness-of-fit testing of six-state Markov model in the Swedish Two-County trial

	Swedish 40-49		Swedish 50-59		Swedish 60-69		Netherlands	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
Prevalence Screen								
DCIS	8	8.06	15	15.08	17	16.93	908	900.55
INV	31	39.95	87	88.69	167	168.96	5548	5640.57
Normal	18456	18446.99	21457	21455.23	20395	20393.11	1077844	1077758.88
First subsequent screen								
DCIS	7	6.01	12	10.67	8	7.85	383	405.13
INV	39	27.01	49	47.26	81	78.04	2223	2059.19
Normal	16396	16405.07	18731	18732.71	16372	16373.94	791790	791879.02
Interval cancer	20.45	24.34	25.28	26.60	37.59	38.74	760.00	812.24
Chi-square (2 df)		8.12		0.33		0.17		19.20
P value		0.0172		0.8488		0.9172		0.00007

For models without interval cancers, the 2-degree of freedom chi-squared figures were 12.01, 1.2 and 1.51 for the UK, Australia, and New York programmes, respectively. The model for the UK (p=0.0025) programme showed a statistically significant lack of fit. (See Table 2.9)

Table 2.9 Model fitting and goodness-of-fit testing of six-state Markov model in the service screening programmes

	UK		Australia		New York	
	Observed	Expected	Observed	Expected	Observed	Expected
Prevalence Screen						
DCIS	2767	2762.87	94	93.47	42	41.13
INV	12323	12412.36	439	440.04	230	231.12
Normal	2520526	2520440.76	75573	75572.49	52378	523.77
First subsequent screen						
DCIS	173	188.64	12	15.66	17	21.74
INV	925	844.42	61	56.50	67	61.89
Normal	227503	226809.25	21433	21413.99	45839	45826.25
No subsequent screen	2291925	2292683.68	54067	54086.85	6455	6468.12
Chi-square (2 df)		12.01		1.24		1.51
P value		0.0025		0.54		0.4703

2.5 Discussion

2.5.1 Clinical Implication

Before considering the implications of our results, we should consider their reliability. The deterministic model tends to yield unstable results, and cannot be regarded as reliable. The assumed Markov process model fits well but not perfectly, for the most part. The estimated rates of progressive and non-progressive DCIS and of progression to invasive disease reflect the raw data, in that programmes with high detection rates of DCIS tend to have high estimated incidence thereof in Table 2.6. Also, the Swedish 60-69 age group, which has the highest detection rate of invasive tumours at prevalence, also has the highest estimated rate of progression from DCIS to invasive disease.

One can never know with certainty what would have happened to a particular case of DCIS if left untreated, and there is a substantial range of uncertainty around the estimated rates of progressive and non-progressive DCIS. However, the estimates are derived from empirical observations on detection of DCIS and invasive breast cancer at screening, and from interval cancer incidence where available. They do not rely on extrapolation of outcome in treated DCIS cases to assumptions about the natural history if treatment had not taken place. Nor do they depend on the representative nature or otherwise of DCIS cases misdiagnosed as benign disease.

On a technical point, the complementary probability in the likelihood function is based on the assumption that the prevalence and incidence screens apply to the same populations. This is approximately true for all studies except New York, for which the two screens are from separate screening programmes. The likelihood approximation seems to work for the New York data, in that estimates compatible with the other programmes are obtained, but it should

be remembered that it is an approximation nevertheless.

The rates of progression in Table 2.6 can be readily transformed to average times spent in the relevant states simply by inversion. Taking the combined estimates at the bottom of the table, the average time a case of screen-detectable progressive DCIS remains in that state before progression to invasive cancer is 1/4.0030 = 0.25 years, or three months. The time is short since it refers to the window of opportunity for early detection, not from DCIS inception to invasion, but from the appearance of the associated calcifications which render it screen-detectable, to the time of invasion. The estimated duration of the preclinical phase once disease has become invasive is 1/0.3990 = 2.51 years, around 30 months. This gives a total combined preclinical phase of 33 months, which is similar to estimates of around 3 years found elsewhere.[89]

Note that the harvest of progressive DCIS is estimated to be very similar at prevalence and incidence screens (Table 2.7). This is consistent with the similar detection rates of DCIS and DCIS: invasive ratios observed in the screening programmes (Table 2.2). It results from the rapid transition to invasive disease once the DCIS becomes calcified and therefore screen-detectable. If there is only 3-4 months to transition to invasive disease, the pool of preclinical tumours at an incidence screen two or three years after the last screen will be much the same as at a prevalence screen. This supports the findings of Evans and colleagues[82] that the majority of screen-detected DCIS is of high grade and that the calcifications whereby the DCIS is detected are a consequence of necrotic debris. An in situ tumour of high grade and with significant necrosis is likely to be at high risk of imminent progression to invasive disease.

Detailed interpretation of the results in Tables 2.6 and 2.7 give rise to some interesting

implications. Table 2.7 indicates that on average 37% of DCIS cases diagnosed at a prevalence screen are non-progressive. The corresponding proportion at an incidence screen is 4%. This is compatible with previous findings that overdiagnosis and length bias are largely phenomena of the prevalence screen.[89] The results suggest that some overtreatment is inevitable but it does not necessarily follow that 37% of all treatment of DCIS cases diagnosed at prevalence screen confers no benefit. In the first place, there is uncertainty about the estimate. In the second, it is not clear that those with non-progressive DCIS have the same risk status as women free of breast carcinoma. Clinical experience suggests that women with breast cancer are at greater risk of new primary breast cancers. If this applies to women with non-progressive DCIS, treatment of these may in some cases forestall future new primaries.

Accepting that there is some diagnosis of non-progressive DCIS and therefore some overtreatment, we can obtain an estimate of the relative burden of this in comparison with the benefit of early treatment of progressive lesions. On the basis of the estimates in Table 2.7, a woman attending for screening for the first time has a 1 in 3,325 chance (30.08 per 100,000) of being diagnosed with a non-progressive DCIS. This is a 19 times smaller chance than the 1 in 175 probability of being diagnosed with a progressive in situ or invasive lesion (519.73 + 51.53 per 100,000). At an incidence screen, the chance of having a non-progressive DCIS lesion diagnosed is 1 in 42,373 (2.36 per 100,000), whereas the chance of having a progressive lesion, whether invasive or in situ, diagnosed is 140 times higher at 1 in 302 (51.60+279.18 per 100,000).

We have specifically avoided using the clinical and pathological aspects of treated DCIS in the above exercise. As stated, this was to avoid the necessity of extrapolation of results in treated DCIS to the estimation of what would happen if the disease were left untreated. It is, however, useful to consider our results in the light of observed behaviour after treatment.

After treatment with local excision alone it is estimated that 18% of DCIS cases recur.[91] Lower rates are observed after mastectomy or after local excision with irradiation.[92] Probability of recurrence is also affected by size and grade of lesion, presence of necrosis and resection margin width in the case of local excision.[93] Opinions vary as to the likely course of disease if left untreated. McCready[92] suggests that progression of untreated DCIS to invasive disease would occur in 25-35% of subjects, whereas Frykberg and Bland suggest that such progression would happen in the majority of cases.[94] Our results are more consistent with the latter. Given the expected detection rates of progressive and non-progressive DCIS in Table 2.7, one would expect 50-80% of DCIS lesions to progress if left untreated. This is also consistent with the results of Evans et al,[82] who found that 61% of screen-detected DCIS cases were of high grade, which is a strong risk factor for progression or recurrence in treated DCIS.[93]

If the results above are accepted, the implication is that the majority of screen-detected DCIS cases would progress if not treated. It is therefore important to target therapy in a way which reflects the risk the lesion poses to the patient. Of course, if a majority of lesions progress, resection is indicated in all cases, but it may be reasonable to reserve mastectomy and use of adjuvant therapies for high-risk lesions. A consolidation of current knowledge to form the basis of practice and further therapeutic research are both indicated.

The implications of our results are therefore:

(1) there is an element of overdiagnosis and overtreatment of DCIS in mammographic screening programmes;

(2) this element is modest compared with the likely benefit of early diagnosis and treatment of progressive lesions; and

(3) increasing diagnosis of DCIS poses a challenge to therapy as much as to early detection.

2.5.2 Methodological implications

One interesting observation is that the stochastic model on the whole performed better than the simpler deterministic model, despite the larger numbers of parameters estimated in the stochastic model. This is likely to be due to the relative wealth of prospective data, which tends to favour the fitting of longitudinal, stochastic models, and to the failure of some of the simplifying assumptions used in the deterministic model.

As already stated, the mover-stayer models above used only detection rates of DCIS, and did not rely on pathological or therapeutic prognostic factors in treated DCIS. Thus our estimates of progression capability and rates are valid estimates of what would have happened in the absence of treatment. It would be of some interest, however, to combine the estimation of mover-styaer model parameters with that of effects on progression in treated DCIS. For this, two compartments of the model might be defined: first a logistic regression model identifying factors predictive of being a mover, i.e. having the propensity of progression to invasive disease in principle; the second an exponential regression or proportional hazards model for time to progression despite treatment, in the movers. One might expect the important predictors in the first compartment to include host factors and pathological aspects of the tumour (for example age, grade of DCIS, size of DCIS), and the second to contain aspects of treatment (such as radiotherapy, mastectomy or local excision, margin status if the latter, and so on). The definition of such a model to preinvasive conditions will be the subject of later chapters.

A final point to note is that this sort of mover-stayer model with latent mover-stayer status is a potentially powerful technique for obtaining empirical estimates of overdiagnosis in disease screening. Up to now, discussion of overdiagnosis has largely centred around assertions as to

the proportion of non-progressive DCIS, which are based on little or no empirical estimation. This approach provides estimates based on actual data. The approach has potential in evaluating screening for premalignant conditions as in cervical, colorectal or oral screening.

Chapter 3 Two-compartment model with covariates—application to adenomas in the bowel

3.1 Introduction

The methods developed in chapter 2 aim to deal with the heterogeneity problem when investigating disease natural history from screening data prior to the effect of treatment. Investigation of heterogeneity may benefit from information on disease progression after treatment, providing a suitable model can be formed. Also, the preceding development did not distinguish the effect of explanatory variables, such as pathological characteristics. A two-compartment model is proposed to investigate the behaviour of movers or stayers in terms of prognostic factors. The first compartment has as the dependent variable the dichotomous potential to progress (or not). The second compartment has as dependent variable the actual progression and observed time to progression for those which have such a potential.

The proposed model is demonstrated in this chapter by the investigation of malignant transformation to colorectal cancer from polyps after polypectomy in terms of size, histological characteristics, and site of the primary lesions. Researchers have demonstrated the importance of histological change and morphologic features of adenomas for the risk of recurrence or progression to cancer.[95-98] However, the statistical methods in those studies were based on traditional regression methods. In this chapter, we estimate the on the probabilities of cases with potential to progress. Data were obtained from a case-cohort study that recruited subjects who underwent their first examination with colonoscopy between 1979 and 1998 in a medical centre in southern Taiwan.

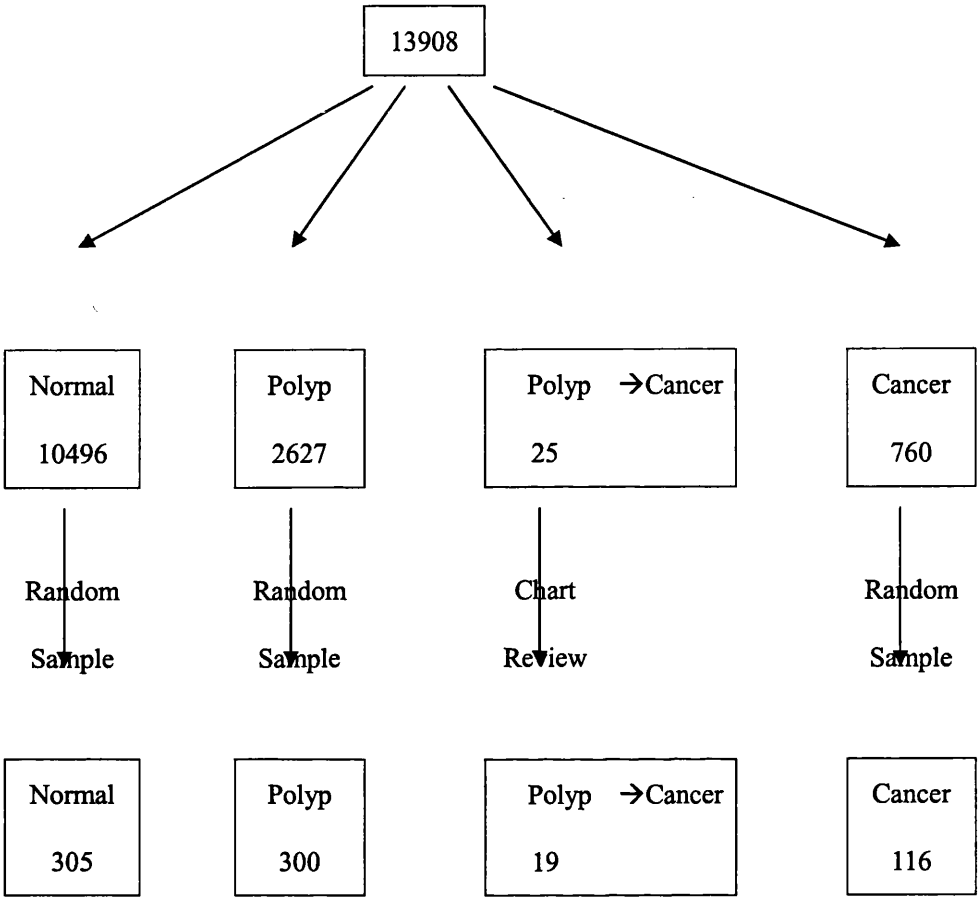
3.2 Materials and Methods

3.2.1 Data

A total of 13908 subjects who underwent the first examination with colonoscopy between 1979 and 1998 in Kaushouing Medical centre, the largest hospital in southern Taiwan, form the cohort. After receiving colonoscopy, this cohort includes three groups, 10496 normal subjects, 2652 patients with polyps, and 760 colorectal cancers. 305 normal subjects and 300 polyp cases that had not yet progressed to invasive carcinoma until the end of 1998 were randomly selected from normal and polyp cohorts, respectively. As regards CRC, a total of 116 cancers were randomly selected. See figure 3.1 for a summary of the design.

All polyps after polypectomy in this cohort were linked to cancer registry data until the end of 1998. A total of 25 CRC cases, regarded as uncensored cases, were identified. These 25 cases plus three hundred polyp cases selected from the above polyp cohort, regarded as censored cases, were used in the present analysis. Patient charts were available for 19 of these 25 cases. Average follow-up for potential malignancy was 3.18 years (3.25 for polyp cases that had not yet progressed, and 2.00 years for malignantly transformed cases).

Figure 3.1 The case-cohort design from Kaushouing Medical centre, Taiwan



3.2.2 Statistical models

Suppose heterogeneity of risk of malignant transition to colorectal cancer exists among patients who received polypectomy, and the pathological characteristics that affect the risk of being potentially progressive can be identified. Then one can expect that the relative risks of observed progression and those of being potentially progressive will be different. The standard logistic regression model was used to model the former, and a mover-stayer model for the latter.

3.2.2.1 Standard model

To investigate the relationship between a binomial event and a set of covariates, one can use the linear regression equation on the logit, or log of the odds:

$$\ln\left(\frac{P_{prog}}{1-P_{prog}}\right)=\sum\beta' X \qquad [3.1]$$

where P_{prog} represents the probability of progression.

So, the probability of progression to colorectal cancer after polypectomy is

$$P_{prog}=\frac{1}{e^{-\sum\beta' X}+1} \qquad [3.2]$$

The full likelihood from n colorectal cancer cases after polypectomy and (N-n) polyps who had not yet progressed to cancer is

$$lik=\left(\frac{1}{e^{-\sum\beta' X}+1}\right)^{n}\left(\frac{e^{-\sum\beta' X}}{e^{-\sum\beta' X}+1}\right)^{N-n} \qquad [3.3]$$

Although it is believed that most colorectal carcinomas develop from preformed adenomas, only a minority of adenomas undergo malignant transformation. An alternative model to handle data with part of the cohort having zero susceptibility is introduced.

In this chapter, we also use the regression models to see how the pathological factors affect the progression rate assuming all polyps have potential to progress. Both semi-parametric methods using the Cox regression model and parametric methods using the accelerated failure time (AFT) model assuming exponential by distributed time to progression are performed.

3.2.2.2 Mover-stayer model

We use another logistic regression formula to represent the relationship between log of the odds of benign polyp with positive susceptibility (mover) and covariates.

$$\ln\left(\frac{P_{mover}}{1-P_{mover}}\right)=\sum\beta' X \qquad [3.4]$$

where P_{mover} is the probability that the polyp can progress.

So, the probability of being a mover is

$$\Pr(\text{Being mover})=\frac{1}{e^{-\sum\beta' X}+1} \qquad [3.5]$$

Therefore, the probability of progression taking heterogeneity into account is the sum of the product of being a mover times the probability of malignant transformation and the product of being a stayer, with zero susceptibility, times zero:

$$
\begin{aligned}
P_1(t_i) &= \Pr(\text{Progress to cancer at } t_i) \\
&= \frac{1}{e^{-\sum \beta' X}+1} \times \lambda e^{-\lambda t_i} + \frac{e^{-\sum \beta' X}}{e^{-\sum \beta' X}+1} \times 0 \\
&= \frac{\lambda e^{-\lambda t_i}}{e^{-\sum \beta' X}+1}
\end{aligned}
\qquad [3.6]
$$

where λ represents the rate of progression if capable of progression. The above assumes an exponential distribution of time to malignant transformation for the movers. In our example, we do not have sufficient event data to model covariates on the actual progression rates.

The probability of no progression is the product of being a mover times the survival function plus the probability of being a stayer:

$$
\begin{aligned}
P_2(t_i) &= \Pr(\text{No progress to cancer by time } t_i) \\
&= \frac{1}{e^{-\sum \beta' X}+1} \times \left(1 - \int_0^{t_i} \lambda e^{-\lambda t_i}\right) + \frac{e^{-\sum \beta' X}}{e^{-\sum \beta' X}+1} \times 1 \\
&= \frac{e^{-\lambda t_i} + e^{-\sum \beta' X}}{e^{-\sum \beta' X}+1}
\end{aligned}
\qquad [3.7]
$$

Under this model, the full likelihood from n colorectal cancer cases after polypectomy and (N-n) polyps who had not yet progressed to malignant cancer becomes

$$
lik = \prod_{i=1}^{n} \left(\frac{\lambda e^{-\lambda t_i}}{e^{-\sum \beta' X}+1} \right) \prod_{i=n+1}^{N} \left(\frac{e^{-\lambda t_i} + e^{-\sum \beta' X}}{e^{-\sum \beta' X}+1} \right) \qquad [3.8]
$$

3.2.3 Estimation taking design into account

Because of the case-cohort design, we use the conditional probability of progression to cancer

and of no progression to cancer given whether the sample was selected (S=1). The corresponding probability formulae are

$$\begin{aligned} P_1^*(t_i) &= \Pr(\text{progress to cancer at } t_i \mid \text{whether to be sampled}) \\ &= \Pr(P_1(t_i) \mid S=1) \\ &= \frac{\Pr(S=1 \mid P_1(t_i)) \times P_1(t_i)}{\sum_{j=1}^{2} \Pr(S=1 \mid P_j(t_i)) \times P_j(t_i)} \\ &= \frac{\pi_1 \times P_1(t_i)}{\sum_{j=1}^{2} \pi_j \times P_j(t_i)}, \text{and} \end{aligned} \qquad [3.9]$$

$$\begin{aligned} P_2^*(t_i) &= \Pr(\text{No progress to cancer by } t_i \mid \text{whether to be sampled}) \\ &= \frac{\pi_1 \times P_1(t_i)}{\sum_{j=1}^{2} \pi_j \times P_j(t_i)}, \end{aligned} \qquad [3.9]$$

where π_1 (=19/25) and π_2 (=300/2652) are random sample fractions for polyp not yet progress to cancer by time t_i and progress to cancer at t_i, respectively.

Therefore, the full likelihood considering sampling fraction based on case-cohort study and the proportion of available data based on missing complete at random assumption from n colorectal cancer cases after polypectomy and (N-n) polyps who had not yet progressed to malignant cancer becomes

$$lik = \prod_{i=1}^{n} P_1^*(t_i) \prod_{i=n+1}^{N} P_2^*(t_i).$$

The Newton-Raphson method in the SAS/IML software was used to estimate maximum likelihood estimates (MLEs) and standard errors.

3.3 Results

We estimate the effects of the predictor variables, size (in categories: less than or equal to 0.5 cm/ greater than 0.5 cm and less than or equal to 1 cm/ greater than 1 cm/ unknown), subsite (section from ascending to transverse colon/ descending colon/ sigmoid and splenic flexure/ rectum and rectosigmoid/ lesions over two or more sites and unknown), and pathology (tubular adenoma/ tubulovillous adenoma/ villous adenoma/ unspecific adenoma and lesions with unknown pathological type). Table 3.1 shows the distribution of the predictor variables in polyp cases which progressed to colorectal cancer and controls.

Table 3.1 Distribution of the predictor variables in progressive cases and controls

	Polyps who had not yet progressed (n=300)	Polyps progressing to CRC (n=19)
Size		
(0 cm, 0.5 cm]	204 (68%)	7 (37%)
(0.5 cm, 1 cm]	28 (9%)	4 (21%)
> 1cm	23 (8%)	5 (26%)
Unknown	45 (15%)	3 (16%)
Site*		
A, C, Hf, T	34 (11%)	1 (6%)
D	18 (6%)	2 (11%)
S, SF	126 (42%)	6 (32%)
R, RS	85 (28%)	9 (47%)
A&D, W,	37 (12%)	1 (5%)
Unknown		
Pathological type**		
T	68 (23%)	8 (42%)
TV	24 (8%)	4 (21%)
V	9 (3%)	2 (11%)
A and others, and Unknown	199 (66%)	5 (26%)

* Subsite abbreviations: A for ascending colon, C for caecum, HF for hepatic flexure, T for transverse colon, D for descending colon, SF for splenic flexure, R for rectum, RS for rectosigmoid, A&D for multiple lesions on ascending and descending colon, W for the whole colon.

** Pathological type abbreviations: T for tubular adenoma, TV for tubulovillous adenoma, V for villous adenoma, and A for nonspecific adenoma.

Table 3.2 shows the odds ratios of progression in the standard logistic model, assuming all adenomas have the potential to progress. We see the highest risk of progression in the lesions with unknown size then a trend of increasing risk with size. The highest risks of progression were observed for subsite rectum or rectosigmoid, and pathological type villous. The table also shows the probability of progression for five hypothetical cases.

Table 3.2 Result of standard logistic regression

Covariate	Beta	SE	OR	Case1	Case2	Case3	Case4	Case5
Intercept	-3.5911	1.02		1	1	1	1	1
Size								
(0 cm, 0.5 cm]	-2.4345	0.93	0.09	0	1	0	0	0
(0.5 cm, 1 cm]	-1.5920	1.06	0.20	0	0	1	0	0
≥ 1cm	-0.6555	0.98	0.52	0	0	0	1	1
Unknown	---	---	1.00	1	0	0	0	0
Site								
A, C, Hf, T	0.4607	1.56	1.59	0	1	0	0	0
D	2.1153	1.48	8.29	0	0	1	0	0
S, SF	1.2574	1.30	3.52	0	0	0	1	0
R, RS	2.3117	1.30	10.09	0	0	0	0	1
A&D, W, Unknown	---	---	1.00	1	0	0	0	0
Pathological type								
T	1.6431	0.69	5.17	0	1	0	0	0
TV	1.4770	0.83	4.38	0	0	1	0	0
V	2.1704	1.03	8.76	0	0	0	1	1
A and others, and Unknown	---	---	1.00	1	0	0	0	0
SCORE				-3.5911	-3.9218	-1.5908	-0.8188	0.2355
Pr(progress)				0.0268	0.0194	0.1693	0.3060	0.5586

Table 3.3 shows the results for the mover-stayer model combined with the logistic regression model for being a mover. The quantitative results of the latter are similar to those of the standard logistic model, but the estimated probabilities of being a mover are rather higher than those of actually progressing. This is because only a subgroup of those adenomas capable of progression actually do so in the time of observation. The rate of progression for the movers is estimated as 0.16, which converts to an average progression time of 6 years. The probability of progression in 6 years is 0.62, which is approximately the rate of the probabilities of potential for progression in Table 3.3 and that of actual progression in Table 3.2, for three out of the five hypothetical examples.

Table 3.3 Result of mover-stayer model

Covariate	Beta	SE	RR	Case1	Case2	Case3	Case4	Case5
Intercept	-3.1492	1.16		1	1	1	1	1
Size								
(0 cm, 0.5 cm]	-2.5788	1.27	0.08	0	1	0	0	0
(0.5 cm, 1 cm]	-1.5339	1.43	0.22	0	0	1	0	0
> 1cm	-0.4215	1.35	0.66	0	0	0	1	1
Unknown	---	---	1.00					
Site								
A, C, Hf, T	0.3073	1.73	1.36	0	1	0	0	0
D	2.2426	1.70	9.42	0	0	1	0	0
S, SF	1.2482	1.49	3.48	0	0	0	1	0
R, RS	2.2420	1.48	9.41	0	0	0	0	1
A&D, W, Unknown	---	---	1.00					
Pathological type								
T	1.8303	0.87	6.24	0	1	0	0	0
TV	1.5158	1.03	4.55	0	0	1	0	0
V	2.2549	1.34	9.53	0	0	0	1	1
A and others, and Unknown	---	---	1.00					
LAMDA	0.1596	0.12						
SCORE				-3.1492	-3.5905	-0.9247	-0.0677	0.9261
Pr(mover)				0.0411	0.0268	0.2840	0.4831	0.7163

3.4 Discussion

The result shown above is from a very preliminary analysis and shows the feasibility of proposed model. There are, however, some concerns about the data.

3.4.1 Missing data

In the present study, pathological characteristics were collected from chart reviewing of randomly sampled data for the non-progressing polyps and for 19 cancers out of 25. For the remaining 6 cancers, patient charts were not available. Among cases with chart available, some pathological characteristics are missing.

3.4.2 Adenomatous type vs Non-adenomatous type

We restricted polyp sampled data to the adenomatous type only. 181 NY (not yet)-progressive polyps (60%) with other histological types (66) or unknown records (115) and 6 CRCs after polypectomy with unknown histological type was excluded from our analysis.

3.4.3 Repeat check-ups between first detection of polyp and diagnosis of cancer

The pathological characteristics used in the present study are the first record of diagnosed polyp. Although there were many repeat colonoscopy check-ups, which may have found other polyps for the patient, corresponding pathological variables were not collected. Therefore, the result here represents the risk from the first diagnosis of first polyp rather than from the temporally closest polypectomy.

3.4.4 Interpretation

There is a further complication of interpretation. The probability of and risk factors for being a mover, and the role of progression are both estimated from data on progression after

polypectomy. As a consequence, it may be that the probabilities of being a mover and the estimated rate of progression for mover are lower bounds on the corresponding probabilities and rate if polypectomy had not occurred.

3.5 Model testing using the simulated data

One hypothetical population with known parameters was created to test

(i) The mathematical accuracy of models,

(ii) The effect of different probability formulae in the analysis of sampled data and of the whole population.

(iii) Whether the mean follow-up time affects the stability of model

(iv) If there exist unobserved variables that can explain some part of heterogeneity, what would be the difference between the true values and estimates from a random-effect model? How to deal with the unobserved heterogeneity? To what extent does ignoring heterogeneity affect estimates of the covariate effects?

The procedure of creating a hypothetical population:

(i) According to the distribution of gender and age in Taiwan, randomly assign the value of sex (0 for female; 1 for male) and age (from 30 to 79 years old) to 10,000 individuals.

(ii) Assume that the distributions of covariates A, B, C follow Bernoulli distributions with parameters 0.06, 0.3, and 0.1, respectively, assign value 0 or 1 for cases without or with the property of that covariate, respectively. Assume there is no dependency among these covariates.

(iii) Create a deterministic relationship between the probability of being a mover and relevant covariates: age, sex, A, B, and C, based on a logistic form,

$$\ln\left(\frac{\Pr_{mover}}{1-\Pr_{mover}}\right) = -5 + 0.6931 * sex + 0.0488 * age + 2.3026 * A + 1.0986 * B + 2.0794 * C$$

(iv) From (iii), the probability of being a mover, $\Pr_{mover}$, can be calculated for each individual.

If random number is less than or equal to Pr_{mover}, then the individual will be labelled as a mover.

(v) For movers, assign their failure times which follow iid exponential distributions with common transition rate 0.2. For stayers, assign their failure times as 99999.

(vi) Randomly assign follow-up times as integers between 1 and 10 years. If the actual failure time derived from (v) for movers is less than or equal to the follow-up time, then the observed survival time is equal to her actual failure time and let CASE (variable represents observed progression)=1. If the actual failure time for movers is greater than follow-up time, i.e. right censored, and for all stayers, the observed survival time equals her follow-up time and let CASE=0.

The SAS code is as follows:

```
data x.hypo;
  do i=1 to 10000;
     sex=rantbl(0,0.4879)-1;
     if sex=1 then do;
      age=rantbl(0,0.0352,0.0352,0.0352,0.0352,0.0352,0.0350,0.0350,0.0350,0.0350,0.0350,
                 0.0316,0.0316,0.0316,0.0316,0.0316,0.0253,0.0253,0.0253,0.0253,0.0253,
                 0.0153,0.0153,0.0153,0.0153,0.0153,0.0149,0.0149,0.0149,0.0149,0.0149,
                 0.0128,0.0128,0.0128,0.0128,0.0128,0.0134,0.0134,0.0134,0.0134,0.0134,
                 0.0105,0.0105,0.0105,0.0105,0.0105,0.0060,0.0060,0.0060,0.0060,0.0060)
                 +29;
      end;
     if sex=0 then do;
      age=rantbl(0,0.0355,0.0355,0.0355,0.0355,0.0355,0.0355,0.0355,0.0355,0.0355,0.0355,
                 0.0322,0.0322,0.0322,0.0322,0.0322,0.0259,0.0259,0.0259,0.0259,0.0259,
                 0.0159,0.0159,0.0159,0.0159,0.0159,0.0160,0.0160,0.0160,0.0160,0.0160,
                 0.0137,0.0137,0.0137,0.0137,0.0137,0.0114,0.0114,0.0114,0.0114,0.0114,
                 0.0084,0.0084,0.0084,0.0084,0.0084,0.0055,0.0055,0.0055,0.0055,0.0055)
                 +29;
      end;
     a=0; if ranuni(0)<=0.06 then a=1;
     b=0; if ranuni(0)<=0.3  then b=1;
     c=0; if ranuni(0)<=0.1  then c=1;

     score=exp(-(-5+0.6931*sex+0.0488*age+2.3026*a+1.0986*b+2.0794*c));

     mover=0; if ranuni(0)<=1/(1+score) then mover=1;
```

```
    if mover=1 then time=ranexp(0)*5;
    if mover=0 then time=99999;

    fu=rantbl(0,0.2,0.2,0.2,0.2);
    if time<=fu then do; case=1; stime=time; end;
    if time> fu then do; case=0; stime=fu;   end;

  output;
  end;
run;
```

Results are shown in Table 3.4 and 3.5. Note that this is the result from only one sampled data set, so slight differences between true values and estimates may result by chance.

Model 1 shows that when the progression rate of movers was fixed as the true value (not exactly the same as the true value of this sample), the corresponding regression coefficients were similar to the true values. After releasing Lamda in Model 2, the regression coefficients were still similar to the true values and a reasonable estimate of lamda was derived. However, Model 2 was unstable when the initial value of lamda was changed. When the initial values are greater than or around the true value, the convergence works well. When the initial value was substantially less than the true value, it converged to some smaller value. (Initial values of 0.15, 0.2, 0.5, 0.95 for lamda converged to around 0.2 and were consistent with each other; initial values of 0.1, 0.05 for lamda converged to around 0.0258). This may mean that the shape of the log-likelihood is not concave, or that the shape is bimodal.

Table 3.4 Results expressed as regression coefficients

Regression Coefficient	True value	Mover-stayer models		Standard model		
		Model 1	Model 2	Logistic regression	Cox regression	AFT model
Intercept	-5.0000	-5.2040	-5.2015	-5.1127	---	-6.4740
Sex	0.6931	0.6525	0.6508	0.5313	0.4961	0.5134
Age	0.0488	0.0546	0.0544	0.0443	0.0376	0.0387
A	2.3026	2.0870	2.0782	1.5621	1.2237	1.2801
B	1.0986	1.0031	1.0000	0.8323	0.6989	0.7179
C	2.0794	2.1144	2.1057	1.6336	1.3438	1.3907
Lamda	0.2000	0.2000 (Fixed)	0.2030	---	---	(0.0015)

Table 3.5 Results expressed as OR's, HR's

Relative Risk	True value	Mover-stayer models		Standard model		
		Model 1	Model 2	Logistic regression	Cox regression	AFT model
Sex	2.00	1.92	1.92	1.70	1.64	1.67
Age	1.05	1.06	1.06	1.05	1.04	1.04
A	10.00	8.06	7.99	4.77	3.40	3.60
B	3.00	2.73	2.72	2.30	2.01	2.05
C	8.00	8.28	8.21	5.12	3.83	4.02

Compared to the results of standard models which did not take heterogeneity into account, we can see large differences between the true value and the estimate of any standard models. Thus, ignoring heterogeneity can lead to substantial bias in the estimates.

In summary, the results showed that when the progression rate (λ) for movers was fixed as the presumed value, the corresponding regression coefficients were similar to their counterparts. After releasing λ to be estimated, the regression coefficients were still stable where initial values for λ were reasonably close to the true value and yield a reasonable estimate of λ. However, the estimation was unstable when the initial value of λ was changed: when the initial values are greater than or around the true value, the convergence works well; when the initial value was substantially less than the true value, it converged to some smaller values.

Compared to the results of standard models which did not take heterogeneity into account, mover-stayer models' performance were superior.

Chapter 4 Disease natural history modelling of adenoma-carcinoma in the large bowel using a multi-state model combined with a frailty model

4.1 Introduction

As noted in previous chapters, in screening for cancer, it is of interest not only whether screening tools can pick up the malignancy in the earliest and most curable preclinical state, but also whether screening can detect a precancerous phase, treatment of which avoid the development of malignancy at all.[99] The success of detecting premalignancy could lead to the reduction of cancer incidence, and, potentially, a corresponding decrease in mortality from cancer. Possible candidates include detection of adenomatous polyps for colorectal cancer,[65;66] leukoplakia or erythroleukoplakia for oral cancer,[100] and squamous intraepithelial lesions for cervical cancer.[101] These possibilities stimulate research on the understanding of disease natural history that would affect policy on screening methodology and interval, and selection of the target population.

The study of disease natural history is, however, difficult because of interruption from treatment immediately after detection of lesions. Follow-up studies of untreated patients or stochastic modelling may be used to elucidate the disease natural history. Stryker et al[102] reviewed 226 patients with untreated colonic polyps greater than or equal to 1 cm in diameter from the retrospective data in Mayo Clinic records from a 6-year period just before the advent of colonoscopy. The cumulative risks of diagnosis of cancer at the site of the index polyp and at any site within the colon at 20-years follow up were found to be 24% and 35%, respectively, in their study. The result, however, may be of doubtful applicability due to the potential limitations of retrospective data and the bias from the selection of patients without treatment. This has a parallel in the studies of biopsy-only-treated cases of ductal carcinoma in situ

(DCIS) of the breast. These are retrospectively identified from pathology archives of cases primarily misdiagnosed as benign diseases. However, such cases are not representative of DCIS in general. In addition, the debate on the two possible pathways for the carcinogenesis of colorectal cancer, 'adenoma-carcinoma sequence' and '*de novo* pathway'[103] (screening is anticipated to detect and treat the premalignancy in the former, and aims to treat frank cancer patients before the clinical syndrome emerges in the latter), complicates the disease natural history. Chen et al[104] used a simple Markov model to address the problem. Annual transition rates were estimated as 0.0095 and 0.022 from adenoma to invasive CRC in their three-state (normal, adenoma, and invasive CRC) model with and without taking the *de novo* pathway into account, respectively. This is consistent in that a higher transition rate from adenoma to cancer should result from the assumption that adenoma provides the only pathway to cancer.

Different mechanisms of disease progression could have major implications for screening policy. In the case that a majority have a very long sojourn time in the detectable and curable stage, screening requires a less frequent interval in terms of economy. However, in the case of considerable heterogeneity, a long interscreening interval is not suitable for everyone. Identification of the group with a high risk of rapidly progressive disease for more frequent surveillance is important in terms of efficiency. Further, in countries with low incidence of colorectal cancer, selective screening with invasive tools such as colonoscopy could be applied to a high-risk group only, instead of offering mass screening to the whole population.

To investigate the heterogeneity, one can use various covariance and regression analyses when all relevant information are available, or use frailty models when dealing with unobservable or unknown factors.[105] Aalen et al[12] used a compound-Poisson-distributed frailty model to interpret the higher incidence of testis cancer among younger men and concluded that the decrease from a certain age was because of a selection phenomenon rather than declining

carcinogenesis with age. There have been various studies using frailty models to address heterogeneity among tumours, but most of them dealt with the traditional survival problem, a two-state rather than multi-state model.[13;15]

For multi-state processes, although simple Markov models have been widely applied to elucidate the natural history of cancer,[106-110] they assume homogeneous transition rates in the population. Combination of a progressive Markov model with constant hazard rates in sequence throughout the process of disease natural history within subjects, with a time-dependent population hazard rate due to heterogeneity between subjects by means of frailty model is the basic idea of the present chapter.

4.2 Objectives

In this chapter, we use data from a case-cohort study which collected results from patients who underwent their first endoscopy examinations in one medical centre in Southern Taiwan to

(1) demonstrate the derivation of the likelihood when introducing Gamma- and compound-Poisson-distributed frailty in a multi-state model;

(2) demonstrate the derivation of the likelihood for a mover-stayer model of a multi-state disease process;

(3) estimate the parameters of the progressive disease natural history model, taking into account heterogeneity of malignant transformation rate from adenoma to colorectal cancer, with and without consideration of adenoma size or histological type;

(4) compare the estimated results, model validation, and predictive clinical performance of the above heterogeneous models with that of the purely homogeneous model.

We begin with a generalised description of a three-state model with transition rates with no distributional assumptions, and expand the model to a k-state process.

4.3 Materials and Methods

4.3.1 Data sources

Patients who had their first endoscopy examinations (including colonosopy and sigmoidoscopy) between 1979 and 1998 in Kaushouing Medical centre in southern Taiwan were enlisted. All records were classified into three categories: no polyp or colorectal cancer (n=10,496), polyp (n=2,652), and colorectal cancer (n=760) with data on the corresponding examination date, but no further data were collected. To estimate the disease natural history of adenoma-carcinoma, one needs information on age at first examination, and the pathological data to distinguish adenoma from non-adenomatous polyps. It is costly and inefficient to review all medical charts. Therefore, a case-cohort study was conducted to review and collect a set of random samples. A total of 305 disease-free subjects, 300 polyp patients, and 116 colorectal cancer cases (all at first endoscopy examination) were extracted.[104] A three-state Markov process and two five-state Markov processes were applied to this case-cohort study to elucidate the disease natural history of adenoma-carcinoma with and without consideration of adenoma size or histological type.[104] Among the sampled data, the average age and proportion of males for disease-free, adenomatous polyps, carcinoma cases were 49.6, 59.3, 60.6, and 48.2%, 59.7%, and 60.3%, respectively. Details of the study design, data sources, construction of likelihood within the case-cohort design in a simple Markov model, and the estimated results from homogeneous models are described elsewhere.[104]

4.3.2 Statistical models

To study heterogeneity in a multi-state progressive process, we develop the transition

probability formulae in a general way and thereafter introduce the frailty factor.

Suppose there is a three-state progressive model as follows.

$$\text{State 1} \xrightarrow{\lambda_1(t)} \text{State 2} \xrightarrow{\lambda_2(t)} \text{State 3}$$

For example, state 1 might be no disease, state 2 adenomatous polyp and state 3 colorectal cancer. $\lambda_1(t)$ represents the instantaneous transition rate from state 1 to state 2 at time t, and $\lambda_2(t)$ represents the transition rate from state 2 to state 3. The "survival" time t from time origin in state 1 and from time origin of state 2 are denoted by $S_1(t)$ and $S_2(t)$, respectively. The probability of staying in state 1 is then simply given by

$$P_{11}(t) = S_1(t). \qquad [4.1]$$

The probability of observing state 2 at t from time origin of state 1 is expressed as

$$P_{12} = \int_0^t S_1(s) \cdot \lambda_1(s) \cdot S_2(t-s) ds. \qquad [4.2]$$

The probability of observing state 3 at t from time origin of state 1 is given by

$$P_{13} = \int_0^t S_1(s) \cdot \lambda_1(s) \int_0^{t-s} S_2(u) \cdot \lambda_2(u) du ds. \qquad [4.3]$$

The above formulae are in a general form that can be applied to any distribution of transition rates, or equivalently, sojourn times. When the transition rates between states are time and

inter-individual homogeneous, it is the simple Markov model. When there is heterogeneity among individuals, a frailty factor may be introduced. In chronic diseases, heterogeneity is often likely, but modelling a multi-state disease process with heterogeneity, especially in more than two transition rates simultaneously, is impractical from a statistical point of view. This is because the matrix solution of the Kolmogorov equations is no longer possible.

To demonstrate how we can incorporate frailty into the above model, we assume that the heterogeneity exists in $\lambda_2(t)$. Although time-homogeneity is assumed throughout the whole process in the same subject, the population hazard from state 2 is still a function of time due to the frailty in the model. This arises because those with larger values of $\lambda_2(t)$ will progress earlier, so that those remaining in state 2 will have on average lower $\lambda_2(t)$ values than the population at the start of observation. The transition rate from state 1 is not time varying and can be expressed as a constant, λ_1.

The probability formula in expression [4.1] is now

$$\begin{aligned} P_{11}(t) &= S_1(t) \\ &= e^{-\lambda_1 t}. \end{aligned} \qquad [4.4]$$

The probability of observing state 2 or state 3 at t from the time origin of state 1 is dependent on the distribution of the frailty factor. For a Gamma-distributed frailty factor,[2] the probabilities can be rewritten as follows.

$$P_{12}(t) = \int_0^t \lambda_1 e^{-\lambda_1 s} \cdot \frac{\theta^\delta}{[\theta + \lambda_{20}(t-s)]^\delta} ds, \text{ and} \qquad [4.5]$$

$$P_{13}(t) = \int_0^t \lambda_1 e^{-\lambda_1 s} \int_0^{t-s} \frac{\delta \cdot \lambda_{20}}{\theta + \lambda_{20} u} \cdot \frac{\theta^{\delta}}{[\theta + \lambda_{20} u]^{\delta}} du ds, \qquad [4.6]$$

where δ and θ are the shape and scale parameter of the Gamma distribution for the frailty factor, respectively. The parameter space is $\delta, \theta > 0$.

When the frailty distribution is in the compound-Poisson type as studied by Aalen,[3;12] the frailty variable is defined by the sum of N independent random variables which are gamma distributed with scale parameter ν and shape parameter η, and N is a random variable which is Poisson distributed with expectation ρ. The parameter space is $\nu, \eta, \rho > 0$. The probabilities of observing state 2 or state 3 at t from the time origin of state 1 can be rewritten as follows.

$$P_{12}(t) = \int_0^t \lambda_1 e^{-\lambda_1 s} \cdot \exp\left\{\rho\left[1 + \frac{\lambda_{20}(t-s)}{\nu}\right]^{-\eta} - \rho\right\} ds, \text{ and} \qquad [4.7]$$

$$P_{13}(t) = \int_0^t \lambda_1 e^{-\lambda_1 s} \int_0^{t-s} \frac{\frac{\rho\eta}{\nu} \cdot \lambda_{20}}{\left[1 + \frac{\lambda_{20} u}{\nu}\right]^{\eta+1}} \cdot \exp\left\{\rho\left[1 + \frac{\lambda_{20} u}{\nu}\right]^{-\eta} - \rho\right\} du ds. \qquad [4.8]$$

An alternative approach to heterogeneity is the mover-stayer model. When the heterogeneity of λ_2 among population is assumed dichotomous, zero-susceptible, and susceptible with a common hazard rate from state 2 to state 3, subjects follow either of the following two possible pathways:

$$\text{(A)} \qquad \text{State 1} \xrightarrow{\lambda_1} \text{State 2} \xrightarrow{\lambda_2} \text{State 3}$$

λ_1

(B) State 1 → State 2

The above two homogeneous disease processes, (A) and (B), follow a three-state Markov chain model and a two-state Markov chain model, respectively. This is called the mover-stayer model.[56] Using formulae developed in Chapter 2, we can obtain the corresponding transition probability matrix. Let $\mathbf{P}_M^A(t)$ and $\mathbf{P}_M^B(t)$ represent the transition probability matrices in a given time interval, t, for pathway (A) and (B), respectively, and P_{mover} as the proportion of people who will follow pathway (A). The probability of staying in state 1 is still

$$\begin{aligned} P_{11}(t) &= S_1(t) \\ &= e^{-\lambda_1 t}. \end{aligned} \qquad [4.9]$$

The probabilities of observing state 2 or state 3 at t from the time origin of state 1 are simply expressed by

$$P_{12}(t) = P_{mover} \times \mathbf{P}_M^A(t)[1,2] + (1 - P_{mover}) \times \mathbf{P}_M^B(t)[1,2], \text{ and} \qquad [4.10]$$

$$P_{13}(t) = P_{mover} \times \mathbf{P}_M^A(t)[1,3]. \qquad [4.11]$$

The above probability formulae will be used in a three-state adenoma-carcinoma model:

Model I. Disease-free → Adenoma → Colorectal cancer.

We now consider the more general case, again dealing with frailty first, then mover-stayer

models. Allowing only one heterogeneous transition rate among individuals can give a general formula. Let us consider a k-state progressive model. Suppose the frailty factor affects the hazard rate from state i (i<k) to state i+1, $\lambda_i(t)$. The probability formulae can then be generalised into three classes:

(i) the probability of observing state j, 1≤j≤i-1, at t from time origin of state 1,

(ii) the probability of observing state i, at t from time origin of state 1, and

(iii) the probability of observing state j, i+1≤j≤k, at t from time origin of state 1.

Because the population hazard is time-homogeneous in case (i), the Markov property holds. We can obtain the transition probabilities using the formulae in Chapter 2.

For case (ii), the probability would be

$$P_{1i}(t) = \int_0^t P_{1,i-1}(s) \cdot \lambda_{i-1} \cdot S_i(t-s) ds. \qquad [4.12]$$

For case (iii), the probability would be

$$P_{1j}(t) = \int_0^t P_{1,i-1}(s) \cdot \lambda_{i-1} \int_0^s [S_i(u) \cdot \lambda_i(u)] \cdot P_{i+1,j}(t-s-u) du ds. \qquad [4.13]$$

Due to the time-homogeneous population hazard from state i+1, the process after state i+1 is a simple Markov process. Therefore, the transition probabilities can be obtained from the formulae in Chapter 2.

Again, to fit the above the multi-state model with dichotomous hazards (mover-stayer model)

from state i to state i+1 among population, the transition probability of observing state j at t from the time origin of state 1 is

$$P_{1j}(t) = \begin{cases} P_{mover} \times \mathbf{P}_M^A(t)[1, j] + (1 - P_{mover}) \times \mathbf{P}_M^B(t)[1, j] & 1 \le j \le i \\ P_{mover} \times \mathbf{P}_M^A(t)[1, j] & i < j \le k \end{cases} \qquad [4.14]$$

where $\mathbf{P}_M^A(t)[1, j]$ and $\mathbf{P}_M^B(t)[1, j]$ represent the transition probability matrices in t for those with positive and zero susceptibility of transition from state 1, respectively.

From the above derivation, two five-state models considering lesion size and histological type were applied to elucidate the disease natural history of the adenoma-carcinoma sequence considering the adenoma size (Model II.A) and histological type (Model II.B).

Model II.A Disease-free → Diminutive Adenoma → Small adenoma → Large Adenoma → Colorectal cancer.

Model II.B Disease-free → Tubular Adenoma → Tubulovillous adenoma → Villous Adenoma → Colorectal cancer.

The above formulae can be expanded to allow regression in which transition from or to any state is possible except for state i and any states with a non-zero instantaneous transition rate to or from state i.

4.3.3 Likelihood derivation

In Model I, as developed in the previous study,[104] the probability for disease-free state at first

examination at age t conditional on being selected can be expressed as:

$$\begin{aligned} P_1(t) &= \Pr(\text{In state 1 at first examination at age t} \mid \text{being sampled}) \\ &= P_{11}(t \mid S=1) \\ &= \frac{\Pr(S=1 \mid 1,t) \times P_{11}(t)}{\sum_{j=1}^{3}\left[P(S=1 \mid j,t) \times P_{1j}(t)\right]} \end{aligned} \qquad [4.15]$$

The sampling fractions for disease free, polyps, and colorectal cancer are 305/10496, 300/2652, and 116/760 (i.e. 0.03, 0.11 and 0.15, respectively). The probabilities for adenoma (P_2) and colorectal cancer (P_3) at first examination at age t given being selected can be derived in the same way.

The likelihood function can then be obtained by

$$\prod_{j=1}^{3}\prod_{i=1}^{n_j} P_j(m_{ij})^{\delta_{ij}} \qquad [4.16]$$

where m_{ij} is the age at first examination of the ith individual in state j, and δ_{ij} is an indicator for the ith individual in state j (j=1, 2, and 3), i.e. δ_{ij}=1 if ith individual is in stage j, and zero otherwise.

The likelihood function for the 5-state model can be derived in a similar way. The sampling fractions for diminutive adenoma, small adenoma, and large adenoma in model II.A, and for tubular adenoma, tubulovillous adenoma, and villous adenoma in model II.B are all assumed to be 300/2652, since random sampling was carried out uniformly to all adenomas without knowledge of type until after the fact of sampling.

The numerical integration in the interactive matrix language (IML) in SAS software was used to obtain the result of integration for deriving the transition probabilities. The matrix language was also used to optimise the likelihood function by the Newton-Raphson method.[111]

4.4 Results

4.4.1 Three-state model

In Model I, the annual incidence rate of adenoma was estimated as 0.0031 when incorporating a Gamma-distributed frailty factor on transition from adenoma to carcinoma (Table 4.1). With the baseline malignant transition rate of adenoma estimated as 0.0804, the scale and shape parameters for the Gamma distribution of frailty factor were 0.3111 and 0.2846. Note the very high variance of λ_{20}.

The estimated result gives the mean of the frailty factor as

$$E(Z)=\frac{\delta}{\theta}=0.9150.$$

Table 4.1 Estimated results from a natural history model for colorectal cancer allowing a frailty distribution of Gamma type for the transition from adenoma to colorectal cancer

Parameter	Estimate	95% CI
λ_1	0.0031	0.0026 ~ 0.0036
λ_{20}	0.0804	0.0000 ~ 4.1328
θ	0.3111	0.0000 ~15.7124
δ	0.2846	0.0000 ~ 1.1971

In this model the average transition rate from adenoma to colorectal cancer is 0.0804x 0.915=0.0736, the product of the basic transition rate times the mean of the frailty factor. However, the distribution of the frailty factor is extremely positive skewed, so that the mean

is not representative (figure 4.1). Percentiles of the frailty factor and the corresponding transition rates and mean sojourn times (MST) are therefore given in Table 4.2. Integrating for each percentile and taking the average, just over 40% of adenomas would progress to cancer within 20 years. About 50% of adenomas have sojourn time longer than 60 years. These cases are very unlikely to have malignant transition in their lifetime.

Figure 4.1 Probability density function of the frailty factor in Model I with Gamma type

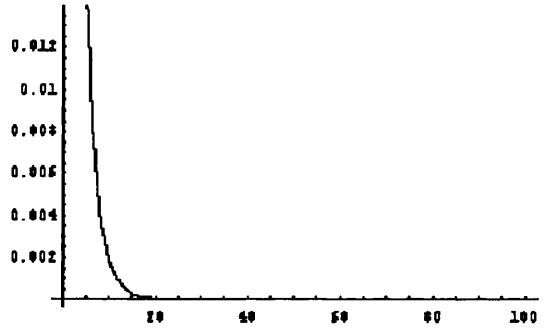


Table 4.2 Distribution of the frailty factor in Model I with Gamma type, the corresponding transition rates and mean sojourn time

Percentile	Frailty factor	λ_2	MST
0th	0	0	∞
10th	0.000680	0.0000547	18288.04
20th	0.007781	0.000626	1598.42
30th	0.032538	0.00262	382.26
40th	0.090669	0.00729	137.18
50th	0.204037	0.01640	60.96
60th	0.405967	0.03264	30.64
70th	0.755622	0.06075	16.46
80th	1.383821	0.1113	8.99
90th	2.714305	0.2182	4.58
100th	∞	∞	0

Similarly, Model I can incorporate frailty factors in the form of the compound Poisson distribution, as in equations [4.7] and [4.8]. The estimated results are shown in Table 4.3. Again, note that there is considerable uncertainty in the estimate of λ_{20}. The annual incidence rate of adenoma was the same as in Table 4.1, 0.0031. A compound Poisson distributed frailty gave the probability of zero susceptibility as

$$\Pr(Z = 0) = \exp(-\rho) = 0.1685 .$$

Given such a model, the expectation of frailty can be obtained as

$$E(Z) = \frac{\rho\eta}{\nu} = 0.4928 .$$

Therefore, the average transition rate from adenoma to colorectal cancer is estimated as 0.1282x0.4928=0.0632.

For those with non-zero susceptibility, the expectation is given as

$$E(Z \mid Z > 0) = \frac{E(Z)}{\Pr(Z > 0)} = 0.5927$$

This means that the malignant transition rate among the susceptible has a mean of 0.0760.

Table 4.3 Estimated results from a natural history model for colorectal cancer allowing a frailty distribution of compound Poisson type for the transition from adenoma to colorectal cancer

Parameter	Estimate	95% CI
λ_1	0.0031	0.0026~ 0.0036
λ_{20}	0.1282	0.0000~ 3.9528
ρ	1.7809	0.0000~12.2086
ν	0.9739	0.0000~33.6948
η	0.2695	0.0000~ 3.0527

Again, for the extremely positive skewed distribution, the mean gives insufficient information. We show the percentiles of the frailty factor in this model in Table 4.4. It implies that some 45% of adenomas would progress to colorectal cancer in 20 years, while the mean sojourn time of 10% of adenomas is less than 5 years. About 50% of adenomas have mean sojourn time longer than 55 years which means that malignant transformation in the lifetime is very unlikely. These results are similar to those using the Gamma frailty model, with only slightly larger probabilities of progression.

Table 4.4 Distribution of the frailty factor in Model I with compound-Poisson type, the corresponding transition rates and mean sojourn time

Percentile	Frailty factor	λ_2	MST
0th	0	0	∞
16.85th	0	0	∞
20th	0.00012	0.000016	63978.47
30th	0.01176	0.001508	663.33
40th	0.05653	0.007247	137.99
50th	0.14408	0.01847	54.14
60th	0.28448	0.03647	27.42
70th	0.49843	0.06390	15.65
80th	0.83652	0.1072	9.32
90th	1.46631	0.1880	5.32
100th	∞	∞	0

The estimated results of Model I by means of a mover-stayer model are shown in Table 4.5. The estimate of annual incidence rate of adenoma, 3.1 per thousand person years, in this model was the same as in models allowing Gamma-distributed and exponentially distributed frailty. The proportion of progressive adenomas was estimated as 55%, in which adenomas progress to malignancy with an estimated hazard of 0.0691 per year. Table 4.5 implies that around 42% of adenomas would progress to cancer within 20 years, after taking the possibility of non-progressive adenomas into account.

Table 4.5 Estimated results from a natural history model for colorectal cancer using a mover-stayer model

Parameter	Estimate	95% CI
λ_1	0.0031	0.0026~0.0036
λ_2	0.0691	0.0000~0.2367
P_{mover}	55.29%	16.04%~94.55%

4.4.2 Five-state model

4.4.2.1 Five-state model for adenoma size

When introducing Gamma-distributed frailty into the hazard rate from diminutive adenoma to small adenoma in Model II.A, the annual incidence rate of diminutive adenoma was estimated as 0.0031. With the baseline transition rate from diminutive adenoma to small adenoma estimated as 0.2632, the scale and shape parameters for the Gamma distribution of the frailty factor were 0.0204 and 0.1696. The annual transition rates from small adenoma to large adenoma and from large adenoma to colorectal cancer were 0.0964 and 0.1462, respectively. The mean sojourn times for small adenomas and large adenomas were 10.37 and 6.84 years, respectively. The Hessian matrix was not negative definitive, however, so the estimate of the variance-covariance matrix could not be obtained.

Table 4.6 Estimated results from a natural history model for colorectal cancer allowing a frailty distribution of Gamma type for the transition from diminutive adenoma to small adenoma

Parameter	Estimate
λ_1	0.0031
λ_{20}	0.2632
λ_3	0.0964
λ_4	0.1462
θ	0.0204
δ	0.1696

The distribution of the frailty factor under this model was again positively skewed (figure 4.2). Table 4.7 implies that about 65% of diminutive adenomas will progress to small adenomas in 20 years. More than 30% of diminutive adenomas would be expected to progress very soon, in one year. About 30% of diminutive adenomas have very long sojourn time, and are likely to have the diminutive adenomas without progression in their lifetime.

Figure 4.2 Probability density function of the frailty factor in Model II.A with Gamma type

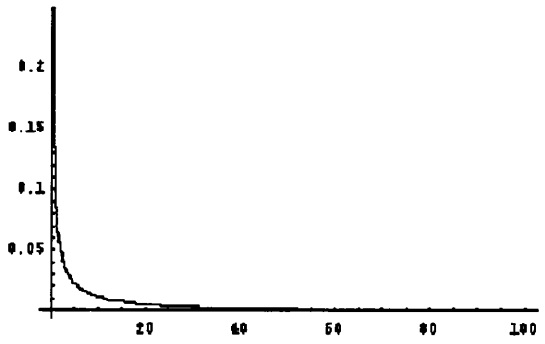


Table 4.7 Distribution of the frailty factor in Model II.A with Gamma-distributed type, the corresponding transition rates and the mean sojourn time

Percentile	Frailty factor	λ_2	MST
0^{th}	0	0	∞
10^{th}	0.000040	0.0000105	95539.16
20^{th}	0.002369	0.000623	1604.02
30th	0.025880	0.00681	146.81
40th	0.141415	0.03722	26.87
50th	0.530690	0.13968	7.16
60th	1.583504	0.41678	2.40
70th	4.102043	1.07966	0.93
80th	9.922485	2.6116	0.38
90th	24.978886	6.5744	0.15
100^{th}	∞	∞	0

Table 4.8 shows the estimated results of Model II.A with compound-Poisson distributed frailty. With this model, variance estimation was possible, although again the baseline rate for the state with frailty has a very high variance. The annual incidence rate of diminutive adenoma was estimated as 0.0031. The characteristics of the compound Poisson distribution gave a probability of zero susceptibility of 6.25%. The annual transition rates from small adenoma to large adenoma and from large adenoma to colorectal cancer were 0.0953 and 0.1450, respectively giving mean sojourn times for small adenomas and large adenomas of 10.49 and 6.90 years, respectively.

Table 4.8 Estimated results from a natural history model for colorectal cancer allowing a frailty distribution of compound Poisson type for the transition rate from diminutive adenoma to small adenoma

Parameter	Estimate	95% CI
λ_1	0.0031	0.0026~ 0.0036
λ_{20}	0.4360	0.0000~30.5312
λ_3	0.0953	0.0473~ 0.1434
λ_4	0.1450	0.0632~ 0.2268
ρ	2.7720	0.0000~30.0007
ν	0.0496	0.0000~ 2.6845
η	0.0809	0.0000~ 1.3619

From the distribution of frailty shown in Table 4.9, about 65% of diminutive adenomas would progress to small adenomas in 20 years. 45% of diminutive adenomas might progress very soon, in one year. About 30% of diminutive adenomas have very long sojourn time, and would be unlikely to progress in the host's lifetime.

Table 4.9 Distribution of the frailty factor in Model II.A with compound-Poisson-distributed type, the corresponding transition rates and the mean sojourn time

Percentile	Frailty factor	λ_2	MST
0th	0	0	∞
6.25th	0	0	∞
20th	0.00028	0.0001217	8218.75
30th	0.01152	0.005025	199.02
40th	0.09419	0.04106	24.35
50th	0.39243	0.1711	5.84
60th	1.15349	0.5029	1.99
70th	2.79486	1.2186	0.82
80th	6.17744	2.6934	0.37
90th	13.99184	6.1004	0.16
100th	∞	∞	0

The estimated results of Model II.A by a mover-stayer model are shown in Table 4.10. The annual incidence rate of diminutive adenoma was consistent with the other models. The transition rate from diminutive adenoma to small adenoma was estimated as 1.6979 among movers, of which the proportion was estimated as 62%. The variation of λ_2 is very large in this model. In movers, the estimated transition rates from small adenoma to large adenoma and from large adenoma to invasive colorectal cancer were 0.0888 and 0.1367, respectively. The results imply that the majority of progressive diminutive adenoma would progress in five years. After then, the remaining diminutive adenomas are mainly non-progressive.

Table 4.10 Estimated results from a natural history model for colorectal cancer considering the lesion size of adenoma using a mover-stayer model

Parameter	Estimate	95% CI
λ_1	0.0031	0.0026~ 0.0036
λ_2	1.6979	0.0000~18.5956
λ_3	0.0888	0.0541~ 0.1234
λ_4	0.1367	0.0695~ 0.2039
P_{mover}	61.86%	52.96%~70.76%

4.4.2.2 Five-state model for histological type

When introducing Gamma-distributed frailty into the hazard rate from tubular adenoma to tubulovillous adenoma in Model II.B, the annual incidence rate of tubular adenoma was estimated as 0.0029 (Table 4.11). With the baseline transition rate from tubular adenoma to tubulovillous adenoma estimated as 0.3236, the scale and shape parameters for the Gamma distribution of the frailty factor were estimated as 0.0239 and 0.1790. The annual transition rates from tubulovillous adenoma to villous adenoma and from villous adenoma to colorectal cancer were 0.0881 and 0.2364, respectively. The mean sojourn times estimated for tubulovillous adenomas and villous adenomas were therefore 11.35 and 4.23 years, respectively.

Table 4.11 Estimated results from a natural history model for colorectal cancer allowing a frailty distribution of Gamma type for the transition rate from tubular adenoma to tubulovillous adenoma

Parameter	Estimate	95% CI
λ_1	0.0029	0.0024~0.0033
λ_{20}	0.3236	0.0000~1.6131
λ_3	0.0881	0.0475~0.1287
λ_4	0.2364	0.0762~0.3966
θ	0.0239	0.0000~0.4620
δ	0.1790	0.0000~0.7561

The distribution for the frailty factor under this model is also positive skewed (figure 4.3). Table 4.12 implies that about 70% of tubular adenomas will progress to tubulovillous adenomas in 20 years. About 40% of tubular adenomas might progress very soon, in one year. Under this model, 30% of tubular adenomas have very long sojourn time. These cases are likely to stay in tubular adenoma without further progression in their lifetime.

Figure 4.3 Probability density function of the frailty factor in Model II.B with Gamma type

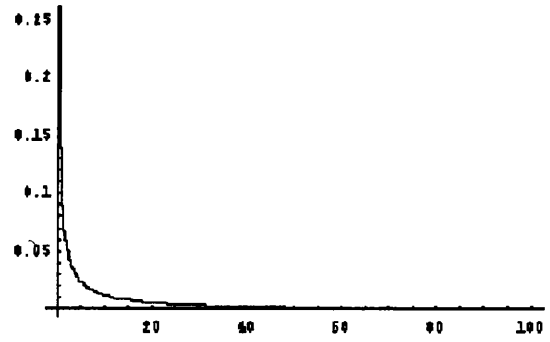


Table 4.12 Distribution of the frailty factor in Model II.B with Gamma-distributed type, the corresponding transition rates and the mean sojourn time

Percentile	Frailty factor	λ_2	MST
0^{th}	0	0	∞
10^{th}	0.0000697	0.0000226	44332.07
20^{th}	0.00335	0.001084	922.46
30^{th}	0.03229	0.01045	95.71
40^{th}	0.1615	0.05226	19.13
50th	0.5664	0.18328	5.4560
60th	1.6014	0.51820	1.9297
70th	3.9703	1.28477	0.7783
80th	9.2598	2.99646	0.3337
90th	22.5807	7.30712	0.1369
100th	∞	∞	0

Table 4.13 shows the estimated results of Model II.B with compound-Poisson distributed frailty. The annual incidence rate of tubular adenoma was estimated as 0.0029. The characteristics of the compound Poisson distribution gave the probability of zero susceptibility as 6.37%. The annual transition rates from tubulovillous adenoma to villous adenoma and from villous adenoma to colorectal cancer were 0.0873 and 0.2346, respectively. With the constant hazard assumption, the mean sojourn times for tubulovillous adenomas and villous adenomas were about 11.45 and 4.26 years, respectively. The Hessian matrix was not negative definitive so that the estimated variance-covariance matrix could not be obtained.

Table 4.13 Estimated results from a natural history model for colorectal cancer allowing a frailty distribution of compound Poisson type for the transition from diminutive adenoma to small adenoma

Parameter	Estimate
λ_1	0.0029
λ_{20}	0.7292
λ_3	0.0873
λ_4	0.2346
ρ	2.7536
ν	0.0942
η	0.0904

Table 4.14 implies that about 65% of tubular adenomas will progress to tubulovillous adenomas in 20 years. 40% of tubular adenomas might progress very soon, in one year. About 30% of tubular adenomas have very long sojourn time, and are likely to have tubular adenoma without progression in their lifetime.

Table 4.14 Distribution of the frailty factor in Model II.B with compound-Poisson-distributed type, the corresponding transition rates and the mean sojourn time

Percentile	Frailty factor	λ_2	MST
0th	0	0	∞
6.37th	0	0	∞
20th	0.00042	0.0003027	3303.25
30th	0.01220	0.008896	112.41
40th	0.08203	0.05982	16.72
50th	0.30003	0.2188	4.57
60th	0.80291	0.5855	1.71
70th	1.81211	1.3214	0.76
80th	3.78560	2.7605	0.36
90th	8.16696	5.9553	0.17
100th	∞	∞	0

The estimated results of Model II.B from a mover-stayer model are shown in Table 4.15. The annual incidence rate of tubular adenoma was consistent with the other models. The transition rate from tubular adenoma to tubulovillous adenoma was estimated as 1.3620 among movers of which the proportion was estimated as 62%. In movers, the estimated transition rates from tubulovillous adenoma to villous adenoma and from villous adenoma to invasive colorectal cancer were 0.0733 and 0.1983, respectively. The results imply that the majority of progressive tubulovillous adenoma would progress in five years. Thereafter, little progression would be observed.

Table 4.15 Estimated results from a natural history model for colorectal cancer considering the lesion size of adenoma using a mover-stayer model

Parameter	Estimate	95% CI
λ_1	0.0031	0.0026~ 0.0036
λ_2	1.6979	0.0000~18.5956
λ_3	0.0888	0.0541~ 0.1234
λ_4	0.1367	0.0695~ 0.2039
P_{mover}	61.86%	52.96%~70.76%

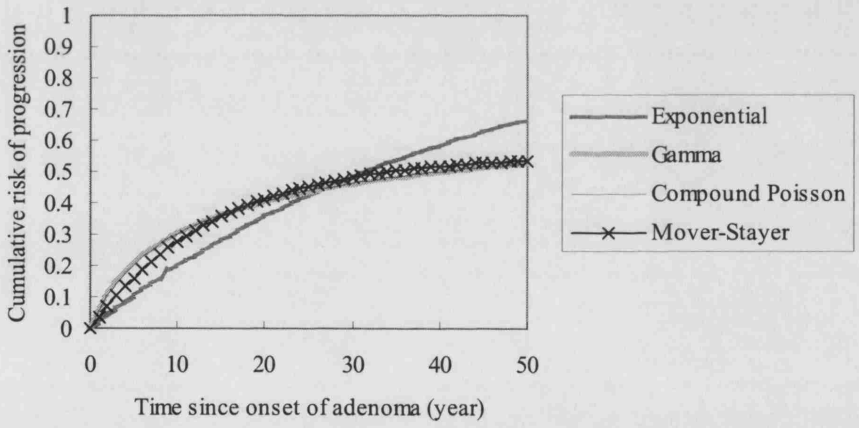
4.4.3 Comparison with purely homogeneous models

Chen et al[104] investigated the disease natural history of colorectal cancer by means of purely homogeneous, simple Markov models. The results corresponding to the Model I in this chapter suggested an estimated transition rate from disease free to adenoma as 0.0031 (0.0026-0.0036) per year, and the annual transition rate from adenoma to invasive CRC was estimated as 0.022 (0.016-0.024).[104] Accordingly, the cumulative risk of malignant transformation from adenoma to invasive CRC under a 3-state Markov model can be illustrated by time (Figure 4.4). With knowledge of population survival from adenomas under gamma- and compound Poisson distributed frailty assumptions and the estimated results from Table 4.1 and Table 4.3, one can demonstrate the cumulative risk under these circumstances, and under the dichotomous hazard assumption in a mover-stayer model from Table 4.5. Estimates are shown in Figure 4.4.

Figure 4.4 shows almost identical prognosis from the two curves taking frailty into account,

and a slight difference from the curve of the mover-stayer model, but a clear difference of the homogeneous model from the other three models. When heterogeneity is allowed, the increase in cumulative risks of progresstion are higher in the earlier years, and then decelerate, because the remaining cases are mainly of lower-susceptibility or even non-susceptible. In the homogeneous model, the trend of cumulative risk is closer to linear.

Figure 4.4 Cumulative risk of malignant transformation from adenoma to colorectal cancer in a three-state model of homogeneous (exponentially distributed sojourn time) and heterogeneous (by introducing gamma- and compound Poisson distributed frailty, and a mover-stayer model) transition rates from adenoma to colorectal cancer



Despite the different appearance of cumulative risk of progression from adenoma between the homogeneous and heterogeneous models, all fitted well (Table 4.16). The heterogeneity models provided only a slight improvement in fit, but probably capably, this was most reached for number of invasive carcinomas.

Table 4.16 The observed number and the expected numbers in a 3-state homogeneous model and three 3-state heterogeneous models with gamma- (Model I-G), compound Poisson distributed frailty (Model I-CP), and dichotomous hazard (Model I-MS) assumption

Mode	Observed number	Expected number			
		Markov model	Model I-G	Model I-CP	Model I-MS
Normal	305	304.06	303.95	303.97	303.96
Adenoma	119	123.46	121.52	121.53	121.64
Invasive CRC	116	112.48	114.52	114.50	114.40
Pearson $\chi^2_{(1)}$		0.2742	0.0750	0.0758	0.0830

In the 5-state homogeneous Markov model considering lesion size,[104] the incidence of diminutive adenoma was estimated as 0.0031 (0.0026-0.0036) per year. The transition rates from diminutive to small adenoma, small to large adenoma, and large adenoma to invasive CRC were estimated as 0.038 (0.030-0.047), 0.13 (0.078-0.18), and 0.19 (0.095-0.28), respectively. Figure 4.5 shows the cumulative risks of progression from diminutive adenomas in the Markov model and the three heterogeneous models. It shows a more remarkable discrepancy between homogeneous and heterogeneous models, compared with the curves shown in Figure 4.4. The cumulative risk of progression for diminutive adenomas in the mover-stayer model increases rapidly in the very early years. The compound Poisson and gamma frailties give almost identical estimates of cumulative risk, with rapid progression in the early years (although not as extreme as in the mover-stayer model). Again, the homogeneous Markov model, by definition, gives a homogeneous incidence over time.

Figure 4.5 Cumulative risk of malignant transformation from diminutive adenoma to small adenoma or more severe status in a five-state model of homogeneous (exponentially distributed sojourn time) and heterogeneous (by introducing gamma- and compound Poisson distributed frailty, and a mover-stayer model) transition rates from diminutive adenoma to small adenoma

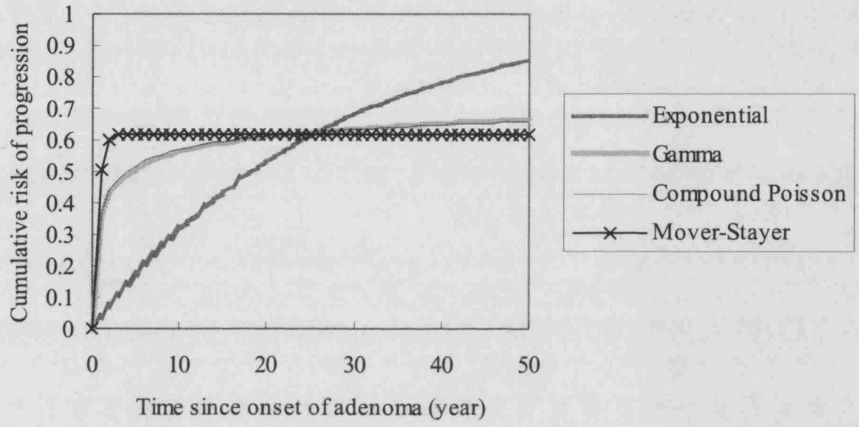


Table 4.17 indicates a good fit for all five-state models taking lesion size into account. The frailty models are only slightly better fitting than the mover-stayer and the homogeneous models.

Table 4.17 The observed number and the expected numbers in a 5-state homogeneous model, three 5-state heterogeneous models with gamma- (Model II.A-G), compound Poisson distributed frailty (Model II.A-CP), and dichotomous hazard (Model II.A-MS) assumption

Mode	Observed number	Expected number			
		Markov model	Model II.A-G	Model II.A-CP	Model II.A-MS
Normal	305	304.26	304.13	304.13	304.14
Diminutive adenoma	78	82.89	79.30	79.19	78.43
Small adenoma	23	22.93	24.63	24.70	25.12
Large adenoma	15	14.68	15.73	15.77	16.01
Invasive CRC	116	112.24	113.21	113.22	113.30
Pearson $\chi^2_{(1)}$		0.4228	0.2345	0.2427	0.3126

In the 5-state homogenous Markov model considering histological type,[104] the incidence of tubular adenoma was estimated as 0.0031 (0.0026-0.0036) per year. The transition rates from tubular to tubulovillous adenoma, tubulovillous to villous adenoma, and villous adenoma to invasive CRC were estimated as 0.038 (0.029-0.047), 0.11 (0.065-0.15), and 0.28 (0.10-0.46), respectively. The comparison of cumulative risks of progression from tubular adenomas, and the fits of the three models, are similar to that of five-state models considering lesion size (Figure 4.6 and Table 4.18). Also, the cumulative risk of progression under all heterogeneity models, but most obviously in the mover-stayer model, shows a sharp increase in the very early years.

Again, the frailty models fit slightly better than the mover-stayer, which in turn fits slightly better than the homogeneous model. All four models, however, provide a good fit.

Figure 4.6 Cumulative risk of malignant transformation from tubular adenoma to tubulovillous adenoma or more severe status in a five-state model of homogeneous (exponentially distributed sojourn time) and heterogeneous (by introducing gamma- and compound Poisson distributed frailty, and a mover-stayer model) transition rates from tubular adenoma to tubulovillous adenoma

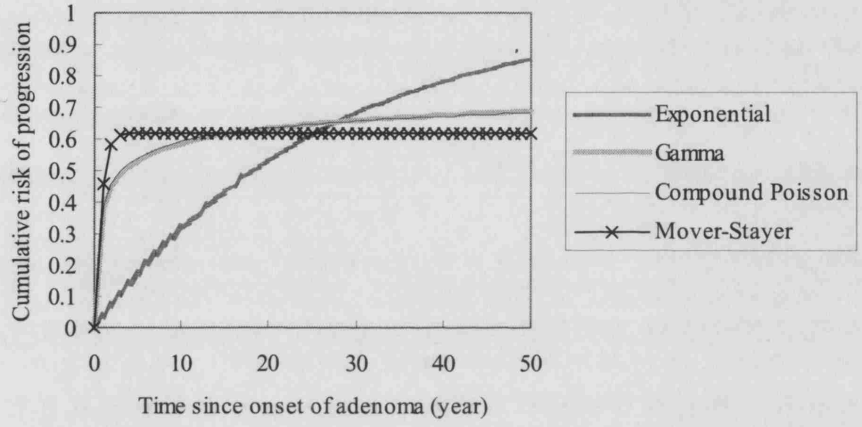


Table 4.18 The observed number and the expected numbers in a 5-state homogeneous model, two 5-state heterogeneous models with gamma- (Model II.B-G) and compound Poisson distributed frailty (Model II.B-CP), and dichotomous hazard (Model II.B-MS) assumption

Mode	Observed number	Expected number			
		Markov model	Model II.B-G	Model II.B-CP	Model II.B-MS
Normal	305	304.33	304.20	304.22	304.16
Tubular adenoma	68	72.35	69.20	69.21	68.43
Tubulovillous adenoma	24	23.94	25.72	25.79	26.29
Villous adenoma	9	8.73	9.43	9.44	9.60
Invasive CRC	116	112.65	113.46	113.37	113.53
Pearson $\chi^2_{(1)}$		0.3710	0.2143	0.2256	0.2953

4.5 Discussion

This chapter incorporated frailty in a general multi-state model and an alternative mover-stayer model to elucidate the disease natural history of the adenoma-carcinoma sequence of colon and rectum. The results were compared with that of a purely homogeneous model in a previous study.[104] The implications of our results are as follows.

4.5.1 Clinical considerations

Stryker et al[102] showed a 24% and a 35% cumulative risk of invasive cancer in the index site and any site in colon, respectively. In our three-state model, the consistent results from heterogeneous models imply that around 40% of adenomas would progress to invasive cancer in 20 years. Although Stryker et al's included large colonic polyps (>1 cm) only, the predicted risk is still lower than ours. This suggests a different casemix in our study from that in Stryker et al.

In the five-state heterogeneous model, a very high early cumulative risk from diminutive or tubular adenomas is seen, particularly in the mover-stayer model. Results suggest a mean sojourn time of less than one year among those who could progress. The majority of those with positive susceptibility would be expected to progress in the early years. This could result from the assumption of only diminutive (or tubular) adenomas having lower or zero susceptibility. The observations of Stryker et al,[102] while quantitatively different from those here are qualitatively consistent with heterogeneity models. The fact that the majority of retrospectively identified adenomas had still not progressed after 20 years strongly suggests a mixture of populations, some with very low or zero susceptibility. The tendency of the cumulative recurrence of treated adenomas to plateau at around ten years is also consistent

with this (Figure 4.7).[95]

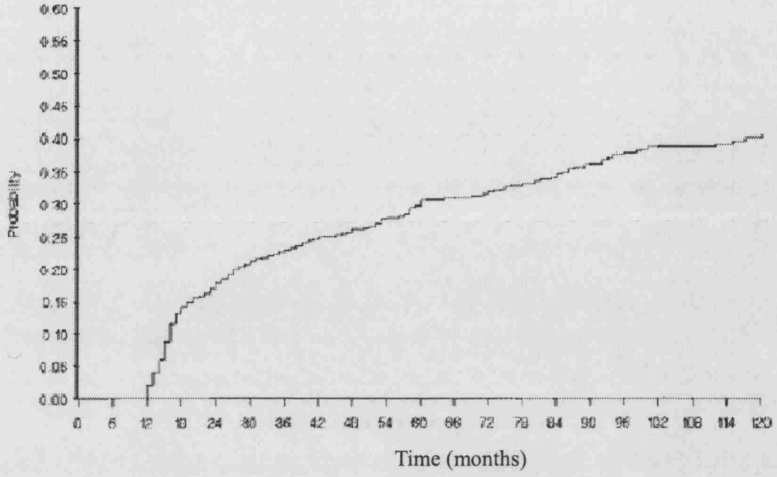


Figure 4.7 Overall cumulative probability of recurrence of adenomatous polyps by time interval after initial polypectomy (Source: Bertario L, et al. (2003) Int. J. Cancer)

4.5.2 Methodology considerations

Estimation was sensitive to the initial values in all heterogeneous models except the mover-stayer model of the three-state disease process. There was similar sensitivity to initial values in the estimation of variances of transition rates in heterogeneity models. This is probably a symptom of the relatively small number of cancer events. Although quite complex longitudinal models can be identifiable in principle, they often need a large body of event data for stable estimation in practice.

With similar good fitting of homogenous and heterogeneous models, one might tend to choose homogenous model in terms of parsimony. Moreover, model selection must be based on a very important issue, practical rationality. From clinical experience, it is unlikely to have extremely different groups in terms of progression from diminutive or tubular adenomas, with

the vast majority of progressive cases progressing in 5 years, as in the results of the 5-state heterogeneous models in this chapter. Also, the possibility of some, although fewer, non-progressive cases among the more severe adenomas suggests inadequacy of the five-state heterogeneous models with heterogeneity in only one transition, as in this chapter.

Two further observations should be noted, however. Firstly, in the three-state model of no disease, adenoma and carcinoma, it is plausible to assume homogeneity of transition to adenoma but heterogeneity of transition from adenoma to cancer. Secondly, the five-state heterogeneity models uniformly give a better fit, albeit only slightly better, than the homogeneous models. In particular, they fit better to the "failures" of the process, the number of invasive carcinoma cases. In addition, the shape of the cumulative incidence cases agrees well with clinical observation of adenoma recurrence.[95] In terms of practicability, fitting frailty models to processes with three states or more is conventionally a daunting prospect. In this chapter, we have demonstrated that if the frailty applies only to a single crucial transition rate between two states, with homogeneous models for transitions between lower and higher states, the model is tractable and estimates can be readily derived with standard software. In theory, this is expandable to more than one transition being subject to heterogeneity. For example, if two transitions were subject to heterogeneity, we would have piecewise homogeneity with three homogeneous processes instead of two. In practice, however, computing and estimation are liable to be difficult. Appendix 1 shows the SAS IML code for the three-state model.

Although the probability formulae derived in this chapter successfully incorporate frailty terms in the multi-state progressive disease process, such methodology has its shortcomings. Firstly, the numerical integration, the only means of estimation due to the lack of closed form of the multiple integration, can be very time-consuming, especially when the number of states

rises. Secondly, the derivation in this chapter can only deal with the progressive model, and does not apply to multiple pathways such as the mixture of adenoma-carcinoma and *de novo* pathways.[104] Finally, for large numbers of states, it is clear that substantial datasets are required for variance estimation. In our example, there was insufficient data for variance estimation for some of the frailty analyses in the five-state models.

Chapter 5 Latent variable analysis in a multi-state disease process for screening data with unobserved or unknown information: the elucidation of the effect of casting-type calcification on the disease process and prognosis of women attending mammographic screening

5.1 Introduction

The situation of unknown or unobserved information is commonly seen in data analysis. Sometimes, this is due to a lack of knowledge of the importance of a particular variable when data are being collected, or sometimes the information is only meaningful or observable for some selective subjects. Calcification features on the mammograms provide such an example. Calcifications are observed as bright dots or lines on a mammogram, usually indicative of calcified necrotic lesions associated with all or part of a tumour which is confined to the milk ducts (ductal carcinoma in situ, DCIS) (Figure 5.1). Calcifications observed in addition to an invasive tumour mass are associated with varying processes depending on the particular radiological appearance of the calcifications. In particular, casting-type calcifications consisting of lines, often branching, are associated with poor prognosis (Figure 5.2). The first study on the role of casting-type calcifications as predictors of long-term survival of small breast cancer, particularly for breast tumours smaller than 15 mm, was published in 2000.[112] Since then, researchers have provided more evidence on the issue.[113-115]

Although many archives of mammograms are potentially available, data on the calcifications were not generally extracted and recorded in the datasets before such features attracted research attention. It is possible to collect data retrospectively from the archive of films. However, this is costly and time-consuming. There may also be problems of loss or inadequate storage of mammograms. The particular interest in calcification features for

invasive breast tumours of size 14 mm or smaller (see below) and the fact that calcifications are often unobservable in larger tumours make smaller tumours the priority of data retrieval. In order to elucidate the multi-state disease process taking the calcification features into account, one needs data from each state, including subjects with normal findings for whom, however, data on calcification features are inapplicable.

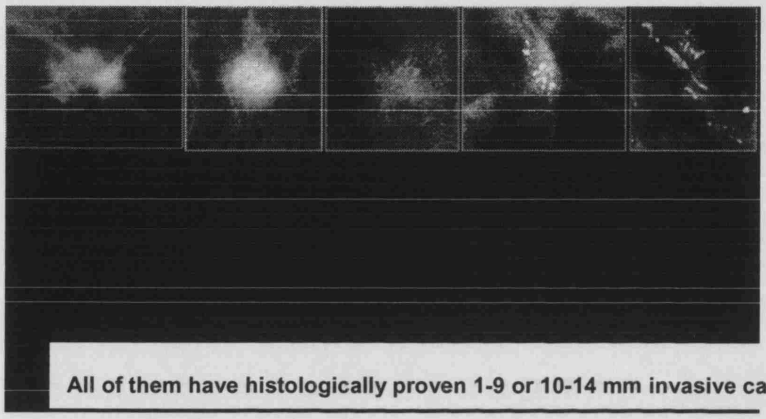


Figure 5.1 Appearance of mammographic features

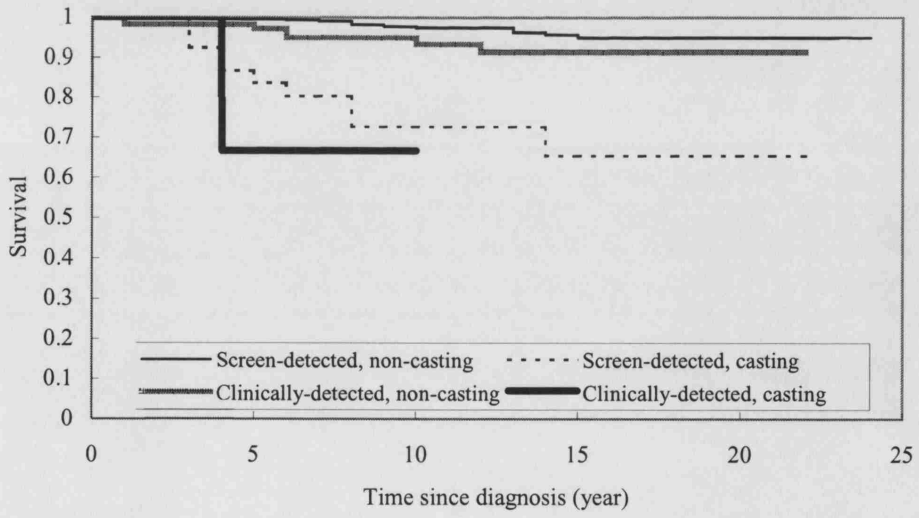


Figure 5.2 Cumulative survival of women with 1-14 mm invasive breast cancer by mammographic appearance and detection mode in Dalarna county, Sweden, 1977-1998

Temporal change in risk factors of interest (or change in their observability) is another issue of concern. For example, it is possible that the calcification features could be ablated over time due to various benign and malignant processes[116] which could cause the situation that a case with casting-type calcifications in the early stage, but not detected until the invasive tumour has grown larger, may not have observable casting-type calcifications by the time of diagnosis, despite still having the corresponding poorer prognosis.

In this chapter, we attempt to fit multi-state models of tumour incidence, tumour size, preclinical/clinical status, and breast cancer death with casting-type calcifications as a covariate on some transitions. Where the presence of casting-type calcifications is unknown or unobservable, it is treated as a latent variable. Such models are compared with another multi-state model taking disease with different characteristics as distinct states and without latent variables involved.

5.2 Objectives

In the present chapter we aimed to use data from the breast cancer mammographic screening programme between 1977 and 1997 in Dalarna county, Sweden

(1) to demonstrate the incorporation of latent class analysis in a multi-state process modelling the disease natural history-prognosis process with inapplicable or unobserved information among subjects in some states,

(2) to quantify the proportion of tumours with casting-type calcifications in the tumour population of size 1-14 mm,

(3) to illustrate the effect of casting-type calcifications on the disease process and prognosis after diagnosis under the assumption that there is no temporal change in casting-type calcifications,

(4) to demonstrate a model which treats the heterogeneity as explicitly different states but without latent variables, and compare the results with those of (1) and (3),

(5) to quantify changes in casting-type calcifications with disease progression from preclinical phase to clinical phase for women with 1-14 mm invasive breast tumours, and

(6) to illustrate the effect of casting-type calcifications on the disease prognosis after diagnosis by detection mode when taking account of the possible temporal change in casting-type calcifications, and compare with the results of (3).

5.3 Material and methods

5.3.1 Data sources

We used mammograhpic screening data from Dalarna county, Sweden, between 1977 and 1997, with two epochs, the randomised trial epoch (1977-1987), in which around 65% of women aged 40-69 were invited to screening, and the service screening epoch (1988-1997), in which all women in this age group were invited.[117] The person-years of attending screening in this period and screening intervals by age groups are listed in Table 5.1. The numbers attending first screening were only available in the trial epoch. The trial epoch included the period during which the control arm received screening at the end of trial.[118] The fact that the coverage rate after 1988 was around 85% suggests that the number of new attendees in the service screening epoch after 1988 was small. Those participants were treated as subsequent screen attendees in our analysis.

Table 5.1 The person-years of attending screening, screening intervals, and the numbers of attendees of the first screen by 5-year age groups in Dalarna county, Sweden, 1977-1997

Age	Person-years		Screening Interval (years)		Numbers attending first
groups	Trial epoch	Service epoch	Trial epoch	Service epoch	screen in the trial epoch
40-44	55 898	87 088	2.00	1.50	7 130
45-49	47 829	87 940	2.00	1.50	7 272
50-54	46 610	75 561	2.75	1.50	8 056
55-59	51 364	65 288	2.75	2.00	9 070
60-64	53 689	65 474	2.75	2.00	8 717
65-69	53 146	68 752	2.75	2.00	8 351

For breast cancer cases, individual data were available on age at diagnosis, detection mode, tumour attributes, including tumour size, and follow-up details until the end of 2001. Data on mammographic appearance of small (<15 mm) invasive breast cancers were extracted from archives in Falun Central Hospital, Sweden. Mammograms of tumours of size 15 mm or more were not classified with respect to calcifications, partly because of prior interest in the small tumours, and partly because clinical experience suggested that calcifications were rare in larger tumours, thus mammographic features including presence or absence of casting-type calcifications, with or without associated tumour mass, were recorded for invasive tumours of size less than 15 mm only. Data were available on 1408 invasive breast cancers. Of these, 1009 (72%) were screen-detected and the remainder interval cancers, i.e. clinically detected. 827 (59%) were of maximum diameter 15 mm or more. All were diagnosed in women aged 40-69. Of the 1409 cases, 212 (15%) died of breast cancer during the maximum 24-year follow-up period.

5.3.2 Statistical models and methods

5.3.2.1 Markov process

A six-state continuous-time Markov process was proposed to model progression of breast cancer in terms of disease status (normal, preclinical detectable phase (PCDP), and clinical phase), tumour size (1-14 mm and >=15 mm), and breast cancer death. The process is depicted in Figure 5.3.

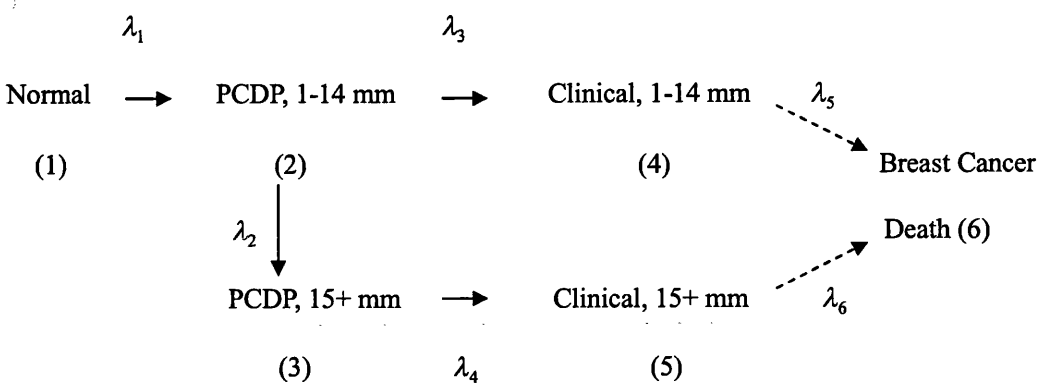


Figure 5.3 The six-state disease process of breast cancer

Accordingly, the transition intensity matrix can be expressed as follows.

$$\mathbf{Q} = \begin{array}{c} \\ \textit{Normal} \\ \textit{PCDP}, 1-14mm \\ \textit{PCDP}, 15+mm \\ \textit{Clinical}, 1-14mm \\ \textit{Clinical}, 15+mm \\ \text{Breast cancer death} \end{array} \begin{array}{c} \begin{array}{cccccc} \textit{Normal} & \textit{PCDP} & \textit{PCDP} & \textit{Clinical} & \textit{Clinical} & \text{Breast cancer} \\ & 1-14mm & 15+mm & 1-14mm & 15+mm & \text{Death} \end{array} \\ \begin{pmatrix} -\lambda_1 & \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & -(\lambda_2+\lambda_3) & \lambda_2 & \lambda_3 & 0 & 0 \\ 0 & 0 & -\lambda_4 & 0 & \lambda_4 & 0 \\ 0 & 0 & 0 & -\lambda_5 & 0 & \lambda_5 \\ 0 & 0 & 0 & 0 & -\lambda_6 & \lambda_6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Again, the transition probabilities can be derived from the transition intensity matrix based on the solution of the Kolmogorov equations[89;90] or by integration, as described in Chapter 2.

Note that transitions among states 1 to 5 follow the disease natural history, and the transitions from state 4 to state 6, and from state 5 to state 6 model the disease prognosis following treatment after clinical detection. Data on survival of screen-detected treated cases cannot be directly modelled by the transition probabilities from state 2 to state 6, or from state 3 to state 6 derived above, because those ignore therapeutic intervention before the onset of clinical stage. Therefore, separate fatality rates, outside of the above process from PCDP, 1-14 mm

(λ_7) and PCDP, 15+ mm (λ_8) to breast cancer death were modelled.

Thus, our combined models do not assume that death from screen-detected disease is impossible; they do, however, rule out direct progression from asymptomatic disease to breast cancer death with no intervening symptoms ever occurring.

Since we are also interested in modelling the effect of casting-type calcifications on the transition rates among small tumours, exponential regression was used for the transition rate from preclinical stage (1-14 mm) to clinical stage (1-14mm) and for the mortality rates from screen-detected (1-14 mm) and from clinical stage (1-14 mm) to breast cancer death. The regression equations are:

$$\lambda_3 = \lambda_{30} \cdot \exp(\beta_3 x)$$
$$\lambda_7 = \lambda_{70} \cdot \exp(\beta_7 x), \text{ and}$$
$$\lambda_5 = \lambda_{50} \cdot \exp(\beta_5 x),$$

where x is the indicator for small tumours with casting-type calcifications. From the above equations, the relative risks (RRs) of dying from breast cancer for cases with casting-type calcifications compared with non-casting cases are the exponential transformations of the $\beta's$.

5.3.2.2 Likelihood construction

In this section, we demonstrate the likelihood derivation by different detection modes prior to the introduction of latent variables. The treatment of latent variables is described in the next section. The transition probabilities can be derived from the transition intensity matrix. Derivation details are given in 5.2.1 and Chapter 2. Let $P_{ij}(m)$ denote the transition

probability from state i to state j in a time interval, m.

5.3.2.2.1 The disease natural history model

At prevalence screen, the probabilities of normal finding, detecting breast tumour smaller than 14 mm, and detecting breast tumour larger than 15 mm are conditional on no clinical disease, as follows:

$$P_{p1.age} = P_{11}(Age)/[P_{11}(Age)+P_{12}(Age)+P_{13}(Age)],$$
$$P_{p2.age} = P_{12}(Age)/[P_{11}(Age)+P_{12}(Age)+P_{13}(Age)], \text{ and}$$
$$P_{p3.age} = P_{13}(Age)/[P_{11}(Age)+P_{12}(Age)+P_{13}(Age)], \text{ respectively.}$$

The conditional probabilities given subjects are in state 1, 2, or 3 at the time of the prevalence screen are used because cases already diagnosed with symptomatic breast cancer are ineligible for the screening programme.

At subsequent screens, the probabilities of normal finding, detecting breast tumour smaller than 14 mm, and detecting breast tumour larger than 15 mm are

$$P_{s1.t} = P_{11}(t),$$
$$P_{s2.t} = P_{12}(t), \text{ and}$$
$$P_{s3.t} = P_{13}(t), \text{ respectively,}$$

where t is the screening interval.

For interval cancers, the probability of breast tumours smaller than 14 mm is

$$P_{ic4.t} = P_{14}(t), \text{ and}$$

the probability of interval cancers larger than 15 mm is

$$P_{ic5.t} = P_{15}(t).$$

Again, t is the screening interval.

5.3.2.2.2 The prognosis model after diagnosis

A patient who dies from breast cancer at time t contributes a term f(t), the density of failure at t , to the likelihood. The contribution from a subject whose survival time is censored at t is $F(t)=1-\int_0^t f(s)ds$, the probability of survival beyond the time of censoring.

The densities of failure time t (for breast cancer deaths) since diagnosis from preclinical breast cancers smaller than 14 mm, from preclinical breast cancers larger than 15 mm, from clinically-detected cases smaller than 14 mm, and from clinically-detected cases larger than 15 mm are

$$f_{s2.t} = \lambda_7 \cdot \exp(-\lambda_7 t),$$

$$f_{s3.t} = \lambda_8 \cdot \exp(-\lambda_8 t),$$

$$f_{s4.t} = \lambda_5 \cdot \exp(-\lambda_5 t), \text{ and}$$

$$f_{s5.t} = \lambda_6 \cdot \exp(-\lambda_6 t), \text{ respectively.}$$

The probabilities of survival at the end of 2001 or death from other causes at time c since diagnosis, for preclinical breast cancers smaller than 14 mm, preclinical breast cancers larger than 15 mm, clinically-detected cases smaller than 14 mm, and clinically-detected cases larger than 15 mm are

$$F_{s2.c} = \exp(-\lambda_7 c),$$

$$F_{s3.c} = \exp(-\lambda_8 c),$$

$$F_{s4.c} = \exp(-\lambda_5 c), \text{ and}$$

$$F_{s5.c} = \exp(-\lambda_6 c), \text{ respectively.}$$

The interactive matrix language (IML) in SAS software version 8.1 was used to compile the code for optimisation for the maximum likelihood estimates (MLEs) and for the second derivatives from which the estimated variance-covariance matrix can be derived.

5.3.2.3 Latent variable analysis

5.3.2.3.1 Model (A): the multi-state process model with a latent variable under the assumption of constant casting-type calcifications

In the previous section, subjects without information on calcifications, such as screen negative subjects, and breast cancer cases with tumours larger than 15 mm, may be considered as a mixture of cases with or without casting-type calcifications which would be observable only in the PCDP at size less than 15 mm. The screen-negative subjects with no detectable disease may be regarded as potentially casting or non-casting cases who have not yet developed disease and who may never do so. The tumours of size 15 mm or more which were not classified with respect to calcifications represent the more straightforward case of an unobserved covariate. To deal with this, we introduce a parameter, p_cast, to represent the proportion of 1-14 mm tumours with casting-type calcifications. Because the temporal change in casting-type calcifications (e.g. from ablation) is not assumed in this model, p_cast denotes the proportion of 1-14 mm tumours with casting-type calcifications in the preclinical state and in the clinical state. Therefore, the probabilities of casting or not in subjects without information on casting-type calcifications can be calculated as a mixture of the corresponding

probabilities of potentially-casting type tumour and of potentially-not-casting type tumour. For example, the probability of detecting 15+ mm breast cancer at subsequent screening with last negative screening result t years ago can be expressed as

$$P_{13}(t) = p\_cast \times P_{13}(t \mid casting) + (1 - p\_cast) \times P_{13}(t \mid non - casting).$$

Three variations of regression model to elucidate the effects of casting-type calcifications on different transition rates were fitted.

Model (A.1): transition from preclinical stage (1-14 mm) to clinical stage (1-14mm) dependent on casting-type calcifications

Model (A.2): transition from preclinical stage (1-14 mm) to clinical stage (1-14mm) and from preclinical stage (1-14 mm) to preclinical stage (15+ mm) dependent on casting-type calcifications

Model (A.3): transition from preclinical stage (1-14 mm) to preclinical stage (15+ mm) dependent on casting-type calcifications

5.3.2.3.2 Model (B): Multi-state model with latent variables under the assumption of temporal change in casting-type calcifications

To deal with the situation that the observed non-casting type small tumours at clinical detection could be a mixture of casting-type calcification tumours in the early PCDP which have since been ablated by natural processes and of non-casting-type tumours in the PCDP, we introduce another parameter, pc_nc, to represent the proportion of cases with casting-type calcifications earlier in their development but whose calcifications were ablated by the time the tumour become palpable and therefore clinically detectable. Taking breast cancer with 1-14 mm tumour detected with clinical symptoms since a last negative screening result t years ago as an example, the corresponding probabilities for observed casting-type and

non-casting-type are

$$P_{14\_c}(t) = p\_cast \times P_{14}(t \mid casting) \times (1 - pc\_nc), \text{ and}$$

$$P_{14\_nc}(t) = p\_cast \times P_{14}(t \mid casting) \times pc\_nc + (1 - p\_cast) \times P_{14}(t \mid non - casting),$$

respectively, where p_cast in Model (B) specifically denotes the proportion of 1-14 mm tumours with casting-type calcifications in the PCDP only.

Note that the newly introduced parameter is highly dependent on the effect of casting-type calcifications on the transition rate from small tumour in the PCDP to small tumour in the clinical phase, β_3. In this model, to avoid problems of collinearity and non-identificability, we abandoned estimation of β_3 and treated it as a constant, 0; that is with a relative risk of unity. This estimates a common transition to clinical disease and therefore a common sojourn time for casting-type and non-casting-type tumours. This assumption is reasonable under the belief that the clinical symptom, usually the awareness of palpable lumps, is independent of calcifications.

5.3.2.4 Model (C): heterogeneity in terms of different states in one multi-state process model

In the previous sections, we treated subjects with different characteristics as different groups that follow multi-state processes with different transition rates. Since the proportion in each group is unobservable, latent variables were applied. In this section, we treat subjects with different characteristics as different states in one multi-state process. By analogy to Model (A), we propose a model of disease progression as follows.

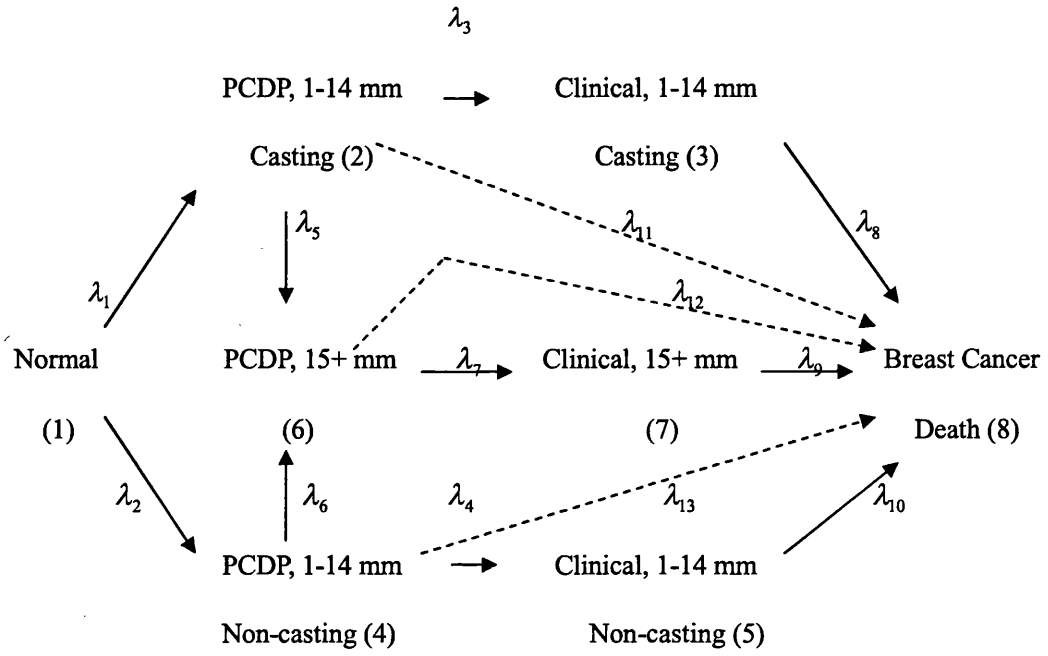


Figure 5.4 An eight-state disease process of breast cancer taking heterogeneity into account by different states

In this model, patients with small breast tumours detected by mammographic screening were classified into two states: with or without casting-type calcifications. If there is no screening interference, each can progress to the clinical phase or can grow to larger than 15 mm prior to clinical symptoms, the state with no information on calcifications. In this model, there is no need to quantify the potential proportion of subjects with normal finding in screenings or in large tumours detected. The derivation of the transition probability matrix, the construction of the likelihood, and the optimisation for the MLEs followed methods described in Chapter 2 and previous sections.

5.4 Results

Results for model (A) are shown in Table 5.2. In Model (A), the estimated annual incidence rates of small preclinical breast cancer were around 2 cases per thousand women years for the three sub-models. The transition rate from preclinical to clinical stage among 1-14 mm tumours unstratified by casting-type calcifications was estimated as 0.1777 in model (A.3). When taking into account the effect of casting-type calcifications on the transition rate from preclinical to clinical stage among 1-14 mm tumours, the transition rate among tumours without casting type calcifications was estimated as 0.1841 in model (A.1) and 0.1829 in model (A.2). The result of exponential regression suggests a statistically non-significant lower risk with casting-type calcifications of progression to clinical disease for small tumours (Risk ratio (RR)=0.39, 95% confidence interval (CI): 0.12-1.27) in model (A.1) and 0.48 (95% CI: 0.13-1.72) in model (A.2)). Casting calcifications were also associated with a non-significantly lower rate of transition from PCDP 1-14 mm to PCDP 15+mm (RR=0.41, 95% CI: 0.07-2.54 in model (A.2) and 0.23 (95% CI: 0.04-1.44) in model (A.3)). The results therefore suggested a greater sojourn time for small tumours with casting-type calcifications. The estimated mean sojourn time for tumours larger than 15 mm in the three models were similar, 1.3 (=1/0.77) years regardless of calcification characteristics. The estimated proportion of observing casting-type calcifications in screen-detected small invasive tumours in model (A.1) was 6.86% (95% CI: 4.74%-8.98%), while in model (A.2) and model (A.3) the estimated proportion were considerably lower, 4.6% and 3.3%, respectively.

Table 5.2 also shows the prognosis model results in terms of breast cancer fatality. The estimated results were almost identical in the three models. Without casting-type calcifications, the fatality rates from breast cancer among women with tumours smaller than 14 mm were 0.0025 and 0.0051 in the PCDP and clinical phase, respectively. For those with

tumours larger than 15 mm (casting data not available), the fatality rates were 0.0168 and 0.0304. Preclinical tumours with casting-type calcifications had 11.27 times the risk (95% CI: 4.94-25.73) of dying of breast cancer in the non-casting type cases, while the risk ratio associated with casting-type calcifications was smaller (RR=8.43, 95% CI: 1.04-68.53) among clinically-detected tumours.

Table 5.3 shows the observed and expected numbers according to the detection mode. The result indicated good model fits in the three models (model (A.1), $\chi^2_{(11)} = 15.36, p = 0.1667$; model (A.2), $\chi^2_{(10)} = 15.84, p = 0.1042$; model (A.3), $\chi^2_{(11)} = 17.46, p = 0.0951$).

Model (A) was also fitted separately in age groups 40-49, 50-59, and 60-69. This led to unstable estimates, particularly for the 40-49 age group. Results are therefore not shown here.

Table 5.2 The estimated results for Model (A)

Parameters	Model (A.1)		Model (A.2)		Model (A.3)	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Natural History component						
Normal → PCDP, 1-14 mm	0.0019	0.0018 ~ 0.0020	0.0019	0.0018 ~ 0.0020	0.0019	0.0018 ~ 0.0020
PCDP, 1-14 mm → PCDP, 15+ mm	1.2421	1.1250 ~ 1.3593				
Non-casting type tumour			1.2953	1.1544 ~ 1.4362	1.3064	1.1756 ~ 1.4373
Reg. Coefficient (Casting/Non-casting)			-0.8891	-2.7101 ~ 0.9318	-1.4528	-3.2717 ~ 0.3661
PCDP, 1-14 mm → Clinical, 1-14 mm					0.1777	0.1416 ~ 0.2139
Non-casting type tumour	0.1841	0.1461 ~ 0.2222	0.1829	0.1450 ~ 0.2207		
Reg. Coefficient (Casting/Non-casting)	-0.9384	-2.1145 ~ 0.2377	-0.7354	-2.0121 ~ 0.5413		
PCDP, 15+ mm → Clinical, 15+ mm	0.7700	0.6855 ~ 0.8545	0.7681	0.6836 ~ 0.8526	0.7662	0.6819 ~ 0.8505
Proportion of casting in PCDP	6.86%	4.74% ~ 8.98%	4.06%	0.38% ~ 7.74%	3.30%	1.22% ~ 5.39%
Prognostic component						
PCDP, 1-14 mm → Breast cancer death						
Non-casting type tumour	0.0025	0.0011 ~ 0.0039	0.0025	0.0011 ~ 0.0039	0.0025	0.0011 ~ 0.0039
Reg. Coefficient (Casting/Non-casting)	2.4224	1.5973 ~ 3.2476	2.4225	1.5980 ~ 3.2469	2.4225	1.5974 ~ 3.2477
Casting-type tumour rate ratio	11.27	4.94 ~ 25.73	11.27	4.94 ~ 25.71	11.27	4.94 ~ 25.73
PCDP, 15+ mm → Breast cancer death	0.0168	0.0134 ~ 0.0202	0.0168	0.0134 ~ 0.0202	0.0168	0.0134 ~ 0.0202
Clinical, 1-14 mm → Breast cancer death						
Non-casting type tumour	0.0051	0.0010 ~ 0.0092	0.0051	0.0010 ~ 0.0092	0.0051	0.0010 ~ 0.0092
Reg. Coefficient (Casting/Non-casting)	2.1312	0.0352 ~ 4.2273	2.1386	0.0246 ~ 4.2526	2.1402	0.0125 ~ 4.2679
Casting-type tumour rate ratio	8.43	1.04 ~ 68.53	8.49	1.02 ~ 70.29	8.50	1.01 ~ 71.37
Clinical, 15+ mm → Breast cancer death	0.0304	0.0241 ~ 0.0367	0.0304	0.0241 ~ 0.0367	0.0304	0.0241 ~ 0.0367

Table 5.3 The goodness of fit for Model (A)

Detection Mode	Tumour size	Casting type	Observed	Expected		
				Model (A.1)	Model (A.2)	Model (A.3)
Prevalence screening						
Normal	N/A	N/A	48407	48422.16	48421.21	48420.91
Breast cancer	1-14 mm	No	82	61.51	61.08	61.30
Breast cancer	1-14 mm	Yes	5	4.91	6.16	6.44
Breast cancer	15+ mm	N/A	102	107.42	107.55	107.35
Subsequent screening						
Normal	N/A	N/A	343739.8	343711.44	343712.85	343713.26
Breast cancer	1-14 mm	No	362	400.50	400.70	402.46
Breast cancer	1-14 mm	Yes	30	31.40	29.39	26.60
Breast cancer	15+ mm	N/A	428	416.48	416.89	417.52
Interval cancer						
Breast cancer	1-14 mm	No	99	95.93	96.03	93.83
Breast cancer	1-14 mm	Yes	3	2.81	2.82	5.02
Breast cancer	15+ mm	N/A	297	279.61	279.94	280.00
Death from Breast cancer						
Screening detected cases						
	1-14 mm	No	13	12.91	12.91	12.91
	1-14 mm	Yes	10	9.51	9.51	9.51
	15+ mm	N/A	92	95.36	95.36	95.36
Interval cancer						
	1-14 mm	No	6	4.92	4.92	4.92
	1-14 mm	Yes	1	1.05	1.05	1.06
	15+ mm	N/A	90	77.91	77.91	77.90
Censored						
Screening detected cases						
	1-14 mm	No	432	431.09	431.09	431.09
	1-14 mm	Yes	25	25.49	25.49	25.49
	15+ mm	N/A	438	434.64	434.64	434.64
Interval cancer						
	1-14 mm	No	93	94.08	94.08	94.08
	1-14 mm	Yes	2	1.95	1.95	1.94
	15+ mm	N/A	207	219.09	219.09	219.10

Table 5.4 shows the estimated results of fitting Model (B). The estimated incidence rate, transition rate from small PCDP to large PCDP, and from large PCDP to large clinical phase were almost identical to those in Model (A.1) (Table 5.2). The average time spent at size 1-14 mm in the PCDP was estimated as 0.7 years $\left(=\frac{1}{1.2420+0.1761}\right)$. The estimated proportion with casting-type calcifications in the small tumours in the PCDP was slightly higher (7.30%, 95% CI: 4.98%-9.63%) than that estimated in Model (A.1). In this model, we also estimated the potential proportion of tumours in which the casting-type calcifications in the PCDP were ablated over time and showed non-calcification when detected with clinical symptoms. The estimated proportion was approximately 60%.

In Model (B), the mortality rates of breast cancer from small screen-detected cases without casting-type calcifications and from large breast cancer were identical with those in Model (A). The small clinically-detected cases without casting-type calcification when they were in the PCDP, had a lower risk of dying from breast cancer compared with their counterparts in Model (A). The latter may have included a proportion of cases who had casting-type calcifications earlier in the disease process. The effects of casting-type calcifications on breast cancer mortality were similar in small screen-detected cases and small clinically-detected cases (RR=11.3 (95% CI: 4.94-25.73), and 11.2 (95% CI: 0.68-185.18), respectively). Table 5.5 shows good model fitting with a Pearson chi-square value of 15.28 on 11 degree of freedom (P=0.1701).

Table 5.4 The estimated results of fitting Model (B)

Parameters	Estimate	95% CI
Natural History Component		
Normal → PCDP, 1-14 mm	0.0019	0.0018 ~ 0.0020
PCDP, 1-14 mm → PCDP, 15+ mm	1.2420	1.1249 ~ 1.3592
PCDP, 1-14 mm → Clinical, 1-14 mm	0.1761	0.1403 ~ 0.2118
PCDP, 15+ mm → Clinical, 15+ mm	0.7700	0.6855 ~ 0.8545
Proportion of casting in PCDP, 1-14 mm	7.30%	4.98% ~ 9.63%
Proportion of transferring to non-casting in Clinical, 1-14 mm	59.69%	25.44% ~ 93.94%
Prognostic Component		
PCDP, 1-14 mm → Breast cancer death		
Non-casting type tumour	0.0025	0.0011 ~ 0.0039
Reg. Coefficient (Casting/Non-casting)	2.4225	1.5973 ~ 3.2476
	11.27	4.94 ~ 25.73
PCDP, 15+ mm → Breast cancer death	0.0168	0.0134 ~ 0.0202
Clinical, 1-14 mm → Breast cancer death		
Non-casting type tumour	0.0037	0.0000* ~ 0.0087
Reg. Coefficient (Casting/Non-casting)	2.4158	-0.3898 ~ 5.2213
	11.20	0.68 ~ 185.18
Clinical, 15+ mm → Breast cancer death	0.0304	0.0241 ~ 0.0367

* Truncated according to the parameter space

Table 5.5 The model fitting of Model (B)

Detection Mode	Tumour size	Casting type	Observed	Expected
Prevalence screening				
Normal	N/A	N/A	48407	48422.18
Breast cancer	1-14 mm	No	82	61.57
Breast cancer	1-14 mm	Yes	5	4.85
Breast cancer	15+ mm	N/A	102	107.41
Subsequent screening				
Normal	N/A	N/A	343740	343711.40
Breast cancer	1-14 mm	No	362	400.39
Breast cancer	1-14 mm	Yes	30	31.55
Breast cancer	15+ mm	N/A	428	416.50
Interval cancer				
Breast cancer	1-14 mm	No	99	95.88
Breast cancer	1-14 mm	Yes	3	2.83
Breast cancer	15+ mm	N/A	297	279.60
Death from Breast cancer				
Screening detected cases				
	1-14 mm	No	13	12.91
	1-14 mm	Yes	10	9.51
	15+ mm	N/A	92	95.36
Interval cancer				
	1-14 mm	No	6	4.98
	1-14 mm	Yes	1	1.03
	15+ mm	N/A	90	77.91
Censored				
Screening detected cases				
	1-14 mm	No	432	431.09
	1-14 mm	Yes	25	25.49
	15+ mm	N/A	438	434.64
Interval cancer				
	1-14 mm	No	93	94.02
	1-14 mm	Yes	2	1.98
	15+ mm	N/A	207	219.09

Table 5.6 shows the estimated results of fitting Model (B) by 10-year age groups. Although more stable than model (A), there were nevertheless problems in estimation, notably in age groups 40-49 and 60-69. The estimated proportion of casting-type calcifications in small tumours in the PCDP was higher in younger age groups (12% for women aged 40-49, 7% for 50-59, and 6% for 60-69). The estimated proportion of tumours in which the casting-type calcifications in the PCDP phase were ablated and showed non-calcification when detected clinically decreased with age, 100% for women aged 40-49, 64% for 50-59, and 16% for 60-69. However, the estimates are extremely imprecise.

Table 5.7 indicated very good model fitting in Model (B) for each 10-year age group sub-model. The Pearson Chi-squares on 9, 11, and 11 degrees of freedom for models with women aged 40-49, 50-59, and 60-69 were 6.09 (p=0.7310), 2.73 (p=0.9938), and 3.60 (p=0.9803), respectively. There are only nine degrees of freedom for the 40-49 age group because it is not applicable for two types of modes, death from breast cancer or censored among interval cancer with casting-type calcifications at size 1-14 mm.

Table 5.6 The estimated results of fitting Model (B) by 10-year age groups

Parameters	40-49		50-59		60-69	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Natural History Component						
Normal → PCDP, 1-14 mm	0.0012	0.0010 ~ 0.0013	0.0019	0.0017 ~ 0.0021	0.0029	0.0027 ~ 0.0031
PCDP, 1-14 mm → PCDP, 15+ mm	2.2337	2.1686 ~ 2.2989	1.2928	1.0659 ~ 1.5198	0.9438	0.8198 ~ 1.0677
PCDP, 1-14 mm → Clinical, 1-14 mm	0.2834	N/A	0.2295	0.1513 ~ 0.3076	0.1233	0.0842 ~ 0.1623
PCDP, 15+ mm → Clinical, 15+ mm	1.0719	N/A	0.8955	0.7163 ~ 1.0747	0.6120	0.5155 ~ 0.7086
Proportion of casting in PCDP, 1-14 mm	12.19%	11.16% ~ 13.23%	6.93%	2.66% ~ 11.19%	6.03%	3.18% ~ 8.88%
Proportion of transferring to non-casting in Clinical, 1-14 mm	100.00%	N/A	63.94%	11.12% ~ 100%*	15.67%	0%* ~ 100%*
Prognostic Component						
PCDP, 1-14 mm → Breast cancer death						
Non-casting type tumour	0.0011	0.000* ~ 0.0027	0.0013	0.000* ~ 0.0030	0.0036	0.0014 ~ 0.0058
Reg. Coefficient (Casting/Non-casting)	1.9937	0.6793 ~ 3.3081	3.0803	1.2903 ~ 4.8703	2.6173	1.6052 ~ 3.6295
	7.34	1.97 ~ 27.33	21.76	3.63 ~ 130.36	13.70	4.98 ~ 37.69
PCDP, 15+ mm → Breast cancer death	0.0225	0.0114 ~ 0.0336	0.0161	0.0097 ~ 0.0226	0.0147	0.0102 ~ 0.0192
Clinical, 1-14 mm → Breast cancer death						
Non-casting type tumour	0.0018	0.000* ~ 0.0059	0.0028	0.000* ~ 0.0123	0.0043	0.000* ~ 0.0103
Reg. Coefficient (Casting/Non-casting)	2.0001	N/A	4.4830	0.7992 ~ 8.1668	-5.9330	N/A
	7.39	N/A	88.50	2.22 ~ 3522.09	0.00	N/A
Clinical, 15+ mm → Breast cancer death	0.0248	0.0113 ~ 0.0382	0.0383	0.0254 ~ 0.0512	0.0290	0.0190 ~ 0.0391

* Truncated according to the parameter space

Table 5.7 The goodness of fit of Model (B) by 10-year age groups

Detection Mode	Tumour size	Casting type	40-49		50-59		60-69	
			Observed	Expected	Observed	Expected	Observed	Expected
Prevalence screening								
Normal	N/A	N/A	14381	14381.30	17078	17074.34	16948	16949.88
Breast cancer	1-14 mm	No	10	5.89	21	19.65	51	43.54
Breast cancer	1-14 mm	Yes	1	0.82	1	1.46	3	2.80
Breast cancer	15+ mm	N/A	10	13.99	26	30.55	66	71.79
Subsequent screening								
Normal	N/A	N/A	153968	153971.53	101287	101290.88	88485	88486.62
Breast cancer	1-14 mm	No	59	61.86	106	108.55	197	203.87
Breast cancer	1-14 mm	Yes	9	8.59	8	8.08	13	13.09
Breast cancer	15+ mm	N/A	111	104.85	118	111.39	199	190.51
Interval cancer								
Breast cancer	1-14 mm	No	25	24.98	36	35.32	38	37.65
Breast cancer	1-14 mm	Yes	0	0.00	1	0.75	2	2.03
Breast cancer	15+ mm	N/A	89	91.47	95	92.78	113	113.29
Death from Breast cancer								
Screening detected cases								
	1-14 mm	No	1	0.88	2	1.86	10	10.51
	1-14 mm	Yes	1	0.87	3	2.36	6	6.94
	15+ mm	N/A	27	26.00	24	24.72	41	44.10
Interval cancer								
	1-14 mm	No	1	0.78	3	2.43	2	1.59
	1-14 mm	Yes	N/A	N/A	1	0.84	0	0.00
	15+ mm	N/A	24	19.53	34	30.22	32	28.46
Censored								
Screening detected cases								
	1-14 mm	No	68	68.12	125	125.14	239	237.49
	1-14 mm	Yes	9	9.13	6	6.64	10	9.06
	15+ mm	N/A	94	95.00	120	119.28	224	220.90
Interval cancer								
	1-14 mm	No	24	24.22	33	33.64	36	36.38
	1-14 mm	Yes	N/A	N/A	0	0.08	2	2.04
	15+ mm	N/A	65	69.47	61	64.78	81	84.54

Table 5.8 shows the estimated result of Model (C). The annual incidence rates of invasive breast tumours with and without casting-type calcifications were estimated as 7.9 per 100,000 and 1.9 per 1000. The overall incidence rate is around 2 per thousand. The estimated transition rates from PCDP 1-14 mm shows that tumours with casting-type calcifications had lower transition rates, except for fatality from breast cancer. The fatality rate of small screen-detected tumours with casting-type calcifications (0.0281) was similar to that of large clinically-detected tumours (0.0304), which in turn was lower than the fatality rate for small clinically-detected tumours with casting-type calcifications (0.0433). Table 5.9 shows a good model fit of Model (C) with a Pearson chi-square value of 15.31 on 10 degrees of freedom (P=0.1210).

Table 5.8 The estimated results of fitting Model (C)

Parameters	Estimate	95% CI
Natural History Component		
Normal → PCDP, 1-14 mm, casting	0.000079	0.000011 ~ 0.000147
Normal → PCDP, 1-14 mm, non-casting	0.0019	0.0017 ~ 0.0020
(casting) PCDP, 1-14 mm → Clinical, 1-14 mm	0.0876	0.000* ~ 0.1949
(non-casting) PCDP, 1-14 mm → Clinical, 1-14 mm	0.1829	0.1451 ~ 0.2207
PCDP, 1-14 mm, casting → PCDP, 15+ mm	0.5353	0.000* ~ 1.4298
PCDP, 1-14 mm, non-casting → PCDP, 15+ mm	1.2951	1.1558 ~ 1.4344
PCDP, 15+ mm → Clinical, 15+ mm	0.7681	0.6836 ~ 0.8526
Prognostic Component		
PCDP, 1-14 mm, casting → Breast cancer death	0.0281	0.0107 ~ 0.0456
PCDP, 1-14 mm, non-casting → Breast cancer death	0.0025	0.0011 ~ 0.0039
PCDP, 15+ mm → Breast cancer death	0.0168	0.0134 ~ 0.0202
Clinical, 1-14 mm, casting → Breast cancer death	0.0433	0.000* ~ 0.1282
Clinical, 1-14 mm, non-casting → Breast cancer death	0.0051	0.0010 ~ 0.0092
Clinical, 15+ mm → Breast cancer death	0.0304	0.0241 ~ 0.0367

* Truncated according to the parameter space

Table 5.9 The model fitting of Model (C)

Detection Mode	Tumour size	Casting type	Observed	Expected
Prevalence screening				
Normal	N/A	N/A	48407	48421.22
Breast cancer	1-14 mm	No	82	61.08
Breast cancer	1-14 mm	Yes	5	6.16
Breast cancer	15+ mm	N/A	102	107.55
Subsequent screening				
Normal	N/A	N/A	343739.8	343328.13
Breast cancer	1-14 mm	No	362	400.24
Breast cancer	1-14 mm	Yes	30	29.37
Breast cancer	15+ mm	N/A	428	416.38
Interval cancer				
Breast cancer	1-14 mm	No	99	96.03
Breast cancer	1-14 mm	Yes	3	2.82
Breast cancer	15+ mm	N/A	297	279.94
Death from Breast cancer				
Screening detected cases				
	1-14 mm	No	13	12.91
	1-14 mm	Yes	10	9.51
	15+ mm	N/A	92	95.36
Interval cancer				
	1-14 mm	No	6	4.92
	1-14 mm	Yes	1	1.05
	15+ mm	N/A	90	77.91
Censored				
Screening detected cases				
	1-14 mm	No	432	431.09
	1-14 mm	Yes	25	25.49
	15+ mm	N/A	438	434.64
Interval cancer				
	1-14 mm	No	93	94.08
	1-14 mm	Yes	2	1.95
	15+ mm	N/A	207	219.09

5.5 Discussion

The present chapter introduced latent variables in the multi-state disease process to deal with unobserved or unknown information in the mammographic screening. The example used was calcification type, which was only known for tumours of size 1-14 mm. The calcification features were found to be a reliable predictor for long-term survival of small invasive breast cancer.[112-115] Using multi-state models, rates of disease progression, disease prognosis in terms of survival, proportions of casting-type calcifications, and proportion of these ablated with time can be simultaneously estimated and adjusted for each other. The results of goodness-of-fit tests suggested the adequacy of the proposed models.

The best fitting model was Model (B), the latent variable model incorporating change in the calcifications over time. The second best fitting model was Model (A.1), the latent variable model with unchanging calcifications status and with regression of the transition rate from preclinical tumours of size 1-14 mm to clinical tumours of size 1-14 mm, on casting-type calcifications.

Another latent variable incorporated in Model (B) was used to model the proportion of casting-type calcification in the PCDP phase which were ablated with time and hence observed as non-casting cases in the clinical phase. However, the problem of unidentibility makes the simultaneous estimation of the proportion of casting-type calcifications ablated and the effect of casting-type calcifications on the instantaneous rate of clinical transformation impossible. In the present study, in Model (B), we assume identity of clinical transition in small tumours between casting- and non-casting-type tumours. Because the clinical symptoms mainly arise as palpable lumps which are an independent way of the appearance of calcifications, it could be argued that the assumption is reasonable.

One of the most striking findings was the observed reduced rate of progression to clinical disease associated with casting-type calcifications in Model (A), in which it is assumed that casting-type calcifications do not change with the tumour's growth. Alternatively, this observation can be qualified by the likelihood that a proportion of clinical symptomatic cancers, which are not observed to have casting-type calcifications, did have these calcifications in the preclinical phase. Model (B) suggests that this proportion is a majority. However, there is considerable uncertainty in estimation of the proportion ablated. Also, clinical impressions suggest that while other calcification types (such as powdery or crushed stone-type calcifications) may be ablated over time, this is not the case for casting-type calcification (Laszlo Tabar, personal communication). On the other hand, it seems unlikely that a feature associated with poor survival should be also associated with slower progression to clinical disease and from small to large tumours.

It is therefore likely that while the models provide a reasonable fit, none of them are biologically accurate. The clinical significant of casting-type calcifications is a recent discovery, and knowledge of the underlying biological process is rudimentary. Better biological models may follow from further pathological research into tumour markers such as hormone receptors and c-erbB/2 status in cases with casting-type calcifications.

The discrepant findings between Model (A) and Model (B) are also illustrated in the disease prognosis in terms of fatality from breast cancer. In Model (A), the risk ratio for dying from breast cancer associated with casting-type calcifications was smaller among clinically-detected tumours than that among screen-detected cases, which is unexpected. Together with the result on longer sojourn time for tumours with casting-type calcifications, it raises the hypothesis of a selection phenomenon whereby the disease process is heterogeneous among cases with casting-type calcification tumours. Although the ideal way to deal with

such a problem is to use a frailty model, insufficient data in the present study led to a failure to fit such a model. In Model (B), under the assumption of possible ablation of casting-type calcifications, the risk ratios for dying from breast cancer associated with casting-type calcifications were similar among screen- and clinically-detected tumours.

The present study results add further support to the findings that casting-type calcifications are associated with poorer prognosis and indicate that absolute mortality in small tumours without calcifications is very small indeed (Figure 5.2). This has two clinical implications. First, there is clearly a need to develop effective treatment regimes for tumours with associated casting-type calcifications. Second, in view of the excellent survival of the group of small tumours without casting-type calcifications, who were treated with surgery alone or surgery plus radiotherapy,[119] the balance of benefits and harms of adjuvant chemotherapy in such cases may contraindicate such therapy.

Age-specific results of both model (A) (not shown) and model (B) (Table 5.6) suggest that some of the parameters change substantially with age. Estimation, however, was imprecise within age subgroups, and variance estimation was not always possible. The proportion of cases with casting-type calcifications reduces with increasing age, and of course it is known that absolute incidence of disease increases with age. Beyond this, any interpretation of the age-specific results must be tentative. Nevertheless, the preliminary results of these tendencies regarding specific age groups provide an important message, that the proportion of casting-type calcifications among small tumours is associated with age. Larger numbers of breast cancer cases for such models will be needed to further quantify age-specific results.

For the proportion of small tumours with casting-type calcifications, we found that the estimate of this proportion was related to the simultaneous estimation of the effect of casting

on the transition from small to large tumours (β_2) in the PCDP, in the three sub-models of Model (A). Without the estimation of β_2 (i.e. assuming it to be zero), the proportion with casting-type calcifications was estimated as around 7% which is slightly smaller than that in Model (B) in which the estimate represented the proportion among small tumours in the PCDP only. Further, with the estimation of β_2, the estimated proportion is much smaller (4% in Model (A.2) and 3.3% in Model (A.3)). This suggests an inadequate estimation of β_2 partly due to collinearity and partly to its estimation depending on numbers of large tumours in the PCDP (for which we do not have information on the presence or not of casting-type calcifications). In addition to the models with latent variables, we also proposed a model to deal with the heterogeneity in terms of different states in one multi-state model (Model (C)). The estimated results yielded similar implications to those of Model (A).

The major limitation of this study is the information on casting-type calcifications which was available from retrospective investigation of invasive tumours with size smaller than 15 mm only. To investigate the entire natural history of tumours with casting-type calcifications and also to examine the hypothesis of frailty mentioned above, we need data not only on small tumours, but also on large tumours and on in situ cases, in which the presence of calcium and necrosis is cytologically proven to be associated with high grade (comedo) cases.[120]

In conclusion, various assumptions were assessed via different multi-state models and demonstrated in this chapter. One assumed that the feature of casting-type calcifications is an unchanged variable in the mammograms, and another assumed that the feature is likely to have temporal change. The results of the first suggest that the progression of tumours with casting-type calcifications is possibly heterogeneous, and those of the second suggest that a high proportion of casting-type calcifications are ablated between the PCDP and the clinical phase. This suggests that mammographic screening may give a good opportunity to catch the

tumours with casting calcifications at an early stage. Clarification and choosing between models need more investigation. We need further investigation with casting data on tumours of all sizes, and further biological studies of histopathological tumour features to clarify these issues. What is clear is that the present study also confirmed with multi-state models the previous findings of survival analyses that the mammographic features represent an important prognostic marker.

Chapter 6 Heterogeneity of abnormalities detected in cervical cancer screening

6.1 Introduction

Cytological screening using Pap smear for cancer of the uterine cervix has been performed in western countries since early 1960s.[121-126] The effect on incidence of and mortality from invasive cervical cancer, although without evidence from randomised controlled trials, has been demonstrated by epidemiological studies. Some 50% mortality reduction from cervical cancer about a decade after the advent of screening was observed in Scandinavia,[123;125;127;128] Iceland,[129] German,[130] Scotland,[131] the USA,[132] Canada,[133] and the UK.[134] In addition to detecting invasive cancer at an early stage, cervical screening also picks up premalignant lesions. Early detection and treatment of such precursor lesions can prevent the incidence of cervical cancer.[135;136] However, the majority of premalignancies would be unlikely to progress to invasive cervical carcinoma and some may regress, [137-139] which makes the traditional methods of estimating the mean sojourn time, such as the difference between modal age in different states or the ratio of prevalence to incidence rate[140], unreliable. Mathematical and statistical models which incorporate the heterogeneity of progression are therefore required to quantify progression or regression sufficiently to inform decisions on frequency of screening.

Mathematical and statistical models are developed for better understanding of the disease natural history and of the effect of different screening regimes.[141] In 1973, Knox[142;143] used a computer macro-simulation technique to fit a comprehensive but complicated model for cervical cancer screening with 26 states, including 'normal', 'reverted normal', 'dysplasia' by regressive or progressive type, 'carcinoma in-situ' by young or old type, 'invasive disease' by occult, early or late stage, 'coned' by disease status, 'treated invasive' by early or late, ……,

'dead of cervical cancer' and 'dead from other causes'. Later, micro-simulation screening analysis (MISCAN) was used to assess cervical cancer screening.[144-146] The macro-simulation used a simulated cohort in which the transitions between states are generated by the application of the assumed probabilities aggregated over the population; the micro-simulation model was one in which individual life histories in a particular population, and the effects of screening on those life histories, were generated by Monte Carlo simulations from stochastic life history and screening models. In simulation modelling procedures, a constellation of parameters with respect to demographic features, epidemiological factors and the disease natural history are assigned initially to simulate a scenario of the observed data. Adjustments of the parameters are iteratively made until the predicted results are close to observed data.

Researchers have also used non-simulation modelling methods to investigate the disease process of cervical cancer. Gustafsson and Adami[147] constructed a five-state disease process including a progression-like and regression-like in situ state in a dynamic model of natural history and screening interference, and calculated parameters by a least squares process comparing the observed and predicted reductions in incidence of and mortality from cervical cancer. Because of limitations of the data source, the preinvasive stage in their model included carcinoma in situ only. Van Oortmarssen and Habbema[148] used a simplified model to test hypotheses about regression against data from the screening programme in British Columbia. Bos and colleagues[149] used a two-step procedure to investigate the non-progression of cervical intraepithelial neoplasia. This incorporated estimation of prevalence and incidence rates of preclinical disease from the observed detection rates, followed by the application to the probability formulae expressed as a function of the proportion of non-progressive cases and the duration of preinvasive stage. Recently, Raffle and colleagues[139] compared the cumulative abnormality rates with numbers expected to develop cancer in the absence of screening to estimate the proportion of high grade dyskaryosis and of high grade dysplasia which would

not progress to cancer.

Other problems in cancer screening include length bias and measurement error. At the prevalence screen, the screen-detected premalignant cases are those with slower progression rates. Therefore the real transition rates of the typical incident cases of dysplasia could be underestimated from the prevalence screen when length-bias exists.

In this chapter, we propose to investigate the disease process of cervical lesions by means of stochastic models, taking account of length bias, measurement error, and heterogeneity. Based on the estimated parameters, a cost-effectiveness analysis is performed, with costs in terms of smears required rather than actual financial sums.

6.2 Objectives

The present chapter aims to establish the disease progression-regression history taking the following items into account:

(1) heterogeneity between lesions detected at prevalence screen and subsequent screens, which is mainly a function of length bias,

(2) the proportion of successfully treated premalignant cases,

(3) measurement error in terms of sensitivity and negative predictive value, and

(4) the heterogeneity of transition rates from pre-malignant lesions to invasive cervical cancer by means of a mover-stayer model.

(5) Finally, we aim to apply the final model to evaluate the cervical cancer screening in terms of additional numbers of smears required to prevent one invasive cervical cancer compared with no screening at all.

6.3 Materials and methods

6.3.1 Data

The cervical screening programme in Taiwan was initiated by the Bureau of National Health Insurance in 1995 and has provided women aged 30 years and over with annual cervical smear and pelvic examination since then.

Between July 1995 and the end of 1998, approximately 2.7 million women were enrolled in this program. Half of them only attended for one screen during this period. The overall coverage for the target population aged over 30 years is 43% and declines with age, ranging from 54% for women aged 30-39 to 24% for women aged 60 and over.

Table 6.1 Frequency of women attending cervical screening programme in Taiwan, 1995-1998

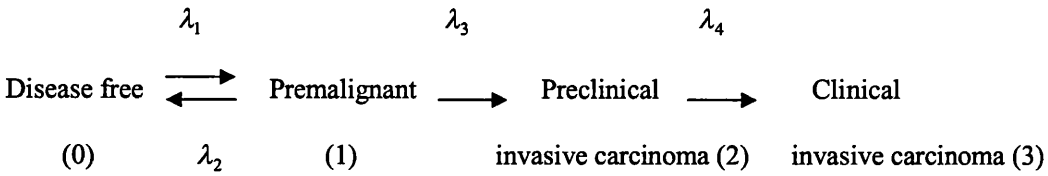
Age	Population at the end of 1998	Women attending screening	Women attending at least two screens
15-29	2,786,815	349,015 (13%)	110,107 (4%)
30-39	1,888,256	1,022,840 (54%)	486,971 (26%)
40-49	1,612,925	735,041 (46%)	335,566 (21%)
50-59	875,964	338,213 (39%)	129,936 (15%)
60-69	680,919	195,599 (29%)	54,360 (8%)
70-79	388,822	72,829 (19%)	14,006 (4%)
80-99	139,630	10,636 (8%)	1,032 (1%)
≥ 15	8,373,331	2,724,173 (33%)	1,131,978 (14%)
≥ 30	5,586,516	2,375,158 (43%)	1,021,871 (18%)

Because our primary interest is in delineating the natural history of cervical carcinoma of the squamous type, we excluded glandular atypia and adenocarcinoma from the primary analysis. Data were obtained from the National Pap Smear Registry. The most severe record was kept if there were duplicate records for the same woman in a single day. All women were linked to the National Cancer Registry to ascertain clinically-detected cancer. Once the date of cancer diagnosis was defined, screening records after that date were deleted.

6.3.2 Models and statistical methods

6.3.2.1 Prior to taking measurement error and heterogeneity of susceptibility of malignant transformation into account

A 4-state Markov process model was used to investigate the disease process of cervical cancer from screening data. The state space S={0; 1; 2; 3} was defined, where 0 stands for the disease-free state and 1, 2, and 3 represent premalignant lesions, preclinical invasive carcinoma and clinical invasive carcinoma, respectively. The premalignant state contains cervical carcinoma in situ, and is allowed to regress. The disease process is depicted as follows.



Let X_{ij} denote transition from state i at previous time to state j at the current observation, where i,j=0,1,2,3. The corresponding probability of transition in a given time interval is denoted as $P_{ij}(t)$. Please refer to Chapter 2 for the derivation of transition probabilities. Let P_{tx}

denote the proportion of successfully treated premalignant lesions, i.e. cases when the lesion is treated and returns to the disease-free state immediately and then follows the disease process from disease-free thereafter. Not all premalignant lesions undergo treatment, and the screening records do not identify treated cases. The individual lesions successfully treated are not statistically distinguishable from the probability of spontaneous regression or from false positives.

To consider the effect of length bias on incidence of premalignant lesions and transition rate from premalignant to preclinical invasive, we introduce the exponential regression form and define

$$\lambda_1 = \lambda_{10} \cdot e^{-\beta_1 x}$$
$$\lambda_3 = \lambda_{30} \cdot e^{-\beta_3 x},$$

where x is an indicator for records from prevalence screen. Length bias is the phenomenon whereby screening detects a disproportionately large number of indolent lesions with very long periods in the preclinical phase. Because of their long preclinical duration, they tend to need only one screen to detect them and therefore they are mainly a phenomenon of the prevalence screen.

At the prevalence screen, all subjects diagnosed with invasive cervical carcinoma prior to screening were excluded from the programme. For subject m receiving screening at age A_m, the probabilities of detecting disease-free, premalignant lesion, and invasive carcinoma at screening are therefore $\frac{P_{00}(A_m)}{P_{00}(A_m)+P_{01}(A_m)+P_{02}(A_m)}$, $\frac{P_{01}(A_m)}{P_{00}(A_m)+P_{01}(A_m)+P_{02}(A_m)}$, and $\frac{P_{02}(A_m)}{P_{00}(A_m)+P_{01}(A_m)+P_{02}(A_m)}$, respectively.

At a subsequent screen, with a time interval T_k from a previous screen with disease-free finding, the outcome could be disease-free, premalignant lesion, and invasive carcinoma with probabilities $P_{00}(T_k)$, $P_{01}(T_k)$ and $P_{02}(T_k)$, respectively. If the previous smear finding is a premalignant lesion, then the transition probability is a mixture of untreated and treated cases. The likelihood of disease-free finding in the subsequent screening with previous premalignant finding is then $(1-P_{tx})\cdot P_{10}(T_k)+P_{tx}\cdot P_{00}(T_k)$. Similarly, probabilities of premalignant and invasive finding are $(1-P_{tx})\cdot P_{11}(T_k)+P_{tx}\cdot P_{01}(T_k)$ and $(1-P_{tx})\cdot P_{12}(T_k)+P_{tx}\cdot P_{02}(T_k)$, respectively.

For invasive carcinoma cases arising due to the appearance of clinical symptoms (interval cancers), the time of clinical invasive cancer onset is known. The likelihood for those with previous disease-free finding after a time interval U_{ic} is $P_{02}(U_{ic})\cdot\lambda_4$. Again, the likelihood for those with previous premalignant finding is the sum of that from untreated or unsuccessfully treated cases, $(1-P_{tx})\cdot P_{12}(U_{ic})\cdot\lambda_4$, and that from treated cases, $P_{tx}\cdot P_{02}(U_{ic})\cdot\lambda_4$.

Until the end of 1998, subjects without any record of screen-detected invasive cancer or clinical-detected invasive cancer (via linkage to the national cancer registry dataset) are taken as censored. For those with previous disease-free and premalignant finding, the likelihood formulae are $(1-P_{03}(V_c))$ and $(1-P_{tx})\cdot[1-P_{13}(V_c)]+P_{tx}\cdot[1-P_{03}(V_c)]$, respectively, where V_c is the interval between last screening and the end of 1998.

6.3.2.2 Measurement error

From the estimated result without taking measurement error and heterogeneity into account, at incidence screen predicted numbers of screen-detected invasive carcinoma cases from

previous disease-free findings are underestimated and numbers from previous premalignant findings are overestimated (see below, Table 6.6 and 6.8). It is presumed that overestimated cases result from possibly false-negative cases, i.e. the sensitivity to premalignant lesions is not 100%. False negatives would imply firstly that the observed previous disease-free group is actually a mixture of disease free and premalignant cases. Secondly, it may also mean that the observed previous premalignant group is not representative of premalignancies as a whole, since those which are misclassified as disease-free may be have different progression rates from the rest of the premalignant population. These unobserved phenomena could lead to a lack of fit in various ways, including that observed. Interval cancer rates after a previous normal finding are similarly underestimated, and are overestimated from a previous premalignant finding. Ignoring sensitivity may lead to the higher estimates of transition rates from premalignant to preclinical invasive carcinoma, and thus cause prediction of more screen-detected cases from premalignant lesions. In this section, therefore, we introduce measurement error into our model.

When taking measurement error into account, test sensitivity (Sen) to premalignant lesion and negative predicted value (NPV) of disease-free from premalignant lesion are considered. The negative predictive value is the probability that a subject classified at screening as disease-free is truly disease-free. It can be expressed as a function of sensitivity, specificity and prevalence, and is used here as it is easier to incorporate into the model than specificity.

At prevalence screening, some screen detected disease-free subjects could be false negative premalignant lesions. The probabilities of detecting disease-free, premalignant lesion, and invasive carcinoma aged A_m become $\frac{P_{00}(A_m)+(1-Sen)\cdot P_{01}(A_m)}{P_{00}(A_m)+P_{01}(A_m)+P_{02}(A_m)}$, $\frac{Sen\cdot P_{01}(A_m)}{P_{00}(A_m)+P_{01}(A_m)+P_{02}(A_m)}$, and $\frac{P_{02}(A_m)}{P_{00}(A_m)+P_{01}(A_m)+P_{02}(A_m)}$, respectively.

At subsequent screening, previous disease-free findings could be composed of truly negative cases (with probability of NPV) or actually premalignant lesion but detected as disease-free (with probability of 1-NPV) in the previous screening. For these cases, the current finding of disease-free could be truly negative or false negative premalignant. Therefore, the probability for cases with observed transition from previous disease-free state to current disease-free state is $NPV \times [P_{00}(T_k) + (1 - Sen) \cdot P_{01}(T_k)] + (1 - NPV) \times [P_{10}(T_k) + (1 - Sen) \cdot P_{11}(T_k)]$. For premalignant findings at current screening with previously observed disease-free, the probability is $NPV \times [Sen \cdot P_{01}(T_k)] + (1 - NPV) \times [Sen \cdot P_{11}(T_k)]$. The probability of currently preclinical invasive finding with previously observed disease-free is $NPV \times P_{02}(T_k) + (1 - NPV) \times P_{12}(T_k)$.

For those having a premalignant finding at the previous screen, the possibility of a false positive is not distinguishable from P_{tx}, but sensitivity still affects the result of the current observation. The population of successfully treated premalignancies contributes to the likelihood. The likelihood formulae for disease-free, premalignant and invasive findings are then $(1 - P_{tx})[P_{10}(T_k) + (1 - Sen) \cdot P_{11}(T_k)] + P_{tx} \cdot [P_{00}(T_k) + (1 - Sen) \cdot P_{01}(T_k)]$, $(1 - P_{tx})[Sen \cdot P_{11}(T_k)] + P_{tx} \cdot [Sen \cdot P_{01}(T_k)]$, and $(1 - P_{tx}) \cdot P_{12} + P_{tx} \cdot P_{02}$, respectively.

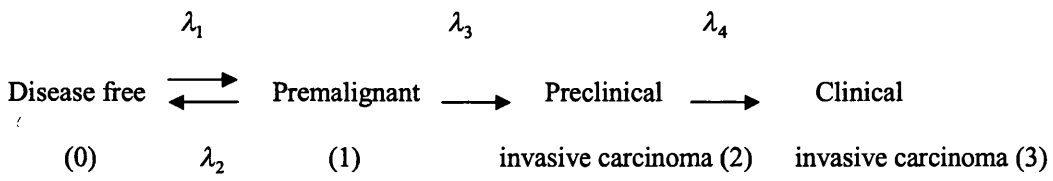
For interval cancers and censored cases, the term of negative predictive value exists in the likelihood terms for those with previous disease-free findings but not for those with previous premalignant findings. Subjects with previous disease-free findings have probabilities of being clinically-detected and censored as $NPV \cdot P_{02}(U_{ic}) \cdot \lambda_4 + (1 - NPV) \cdot P_{12}(U_{ic}) \cdot \lambda_4$ and $NPV \cdot (1 - P_{03}(U_{ic})) + (1 - NPV) \cdot (1 - P_{13}(U_{ic}))$. The corresponding probabilities for those with previous premalignant finding are $(1 - P_{tx}) \cdot P_{12}(U_{ic}) \cdot \lambda_4 + P_{tx} \cdot P_{02}(U_{ic}) \cdot \lambda_4$ and $(1 - P_{tx}) \cdot [1 - P_{13}(U_{ic})] + P_{tx} \cdot [1 - P_{03}(U_{ic})]$.

In the analysis taking measurement error into account, only data from the subsequent screens were used, to avoid the need to estimate length bias effects simultaneously.

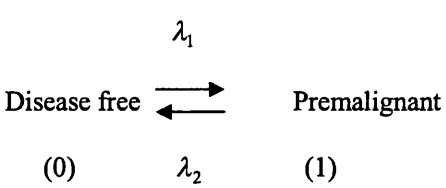
6.3.2.3 Mover-stayer model

As stated above, the majority of premalignant lesions are unlikely to progress to invasive cervical carcinoma. We propose a mover-stayer model in which a proportion, say P_mover, of subjects can follow the entire natural history of cervical cancer from normal to invasive carcinoma, while the rest (1-P_mover) can only progress to premalignanct lesions and then either stay in that state or regress to normal. Within these two groups, the transition rates among subjects are assumed to be homogeneous. The possible pathways of these two groups are depicted as follows.

Model (A) (for mover)



and Model (B) (for stayer)



Under the assumption of homogeneity in the submodels, the above two submodels follow a four-state and a two-state Markov model, respectively. The transition probability matrixes in model (A) and model (B) are represented by $\mathbf{P}_M^A(t)$ and $\mathbf{P}_M^B(t)$ using the formulae derived

in Chapter 2. Therefore, the transition probability matrix in section 6.3.2.1 and 6.3.2.2 becomes

$$\begin{cases} \mathbf{P}(t)[i,] = P_{mover} \times \mathbf{P}_M^A(t)[i,] + (1 - P_{mover}) \times \mathbf{P}_M^B(t)[i,] & \text{for i} = 1,2 \\ \mathbf{P}(t)[i,] = \mathbf{P}_M^A(t)[i,] & \text{for i} = 3,4 \end{cases}$$

6.3.3 Computer simulation

After estimating the progression and screening parameters from a credible and well-fitting model, there remains the problem of determining the implications of these for the screening regime. To address this problem, we used computer simulation, using the parameter estimates of a well-fitting model to generate premalignancies, preclinical invasive cases, clinical cancers, and screening-detected events. Cohorts of 1,000 women aged 30, 40, 50, and 60 years were simulated until the age of 79 years in five arms with screening intervals of 1, 3, 5 and 10 years, and for no screening at all. We estimated the outcomes in those actually attending, and assumed sensitivities of 95% and 90% to the premalignant lesions, specificity of 98%, and successfully treated proportions of 100% and 85%. These two sensitivities and single specificity were consistent with observed results (see below). Two scenarios were simulated: no screening before the start age (the eligible pool includes subjects in disease-free, preclinical, and clinical state at entry); and population already intensively screened (all women start from normal state). The former case corresponds to screening starting at the lower point of each age group considered, so that, for example, the results for age group 40-79 correspond to the benefit of starting screening at age 40 and ending at age 79. The latter pertains to screening in a given age group assuming all prevalent preclinical disease is already excluded by previous screening. Thus the results for age group 40-79 would then correspond to the benefit of continuous screening for age 40 to age 79 given that intensive screening had been performed before age 40. Figure 6.1 shows the flowchart of the computer simulation in the first scenario.

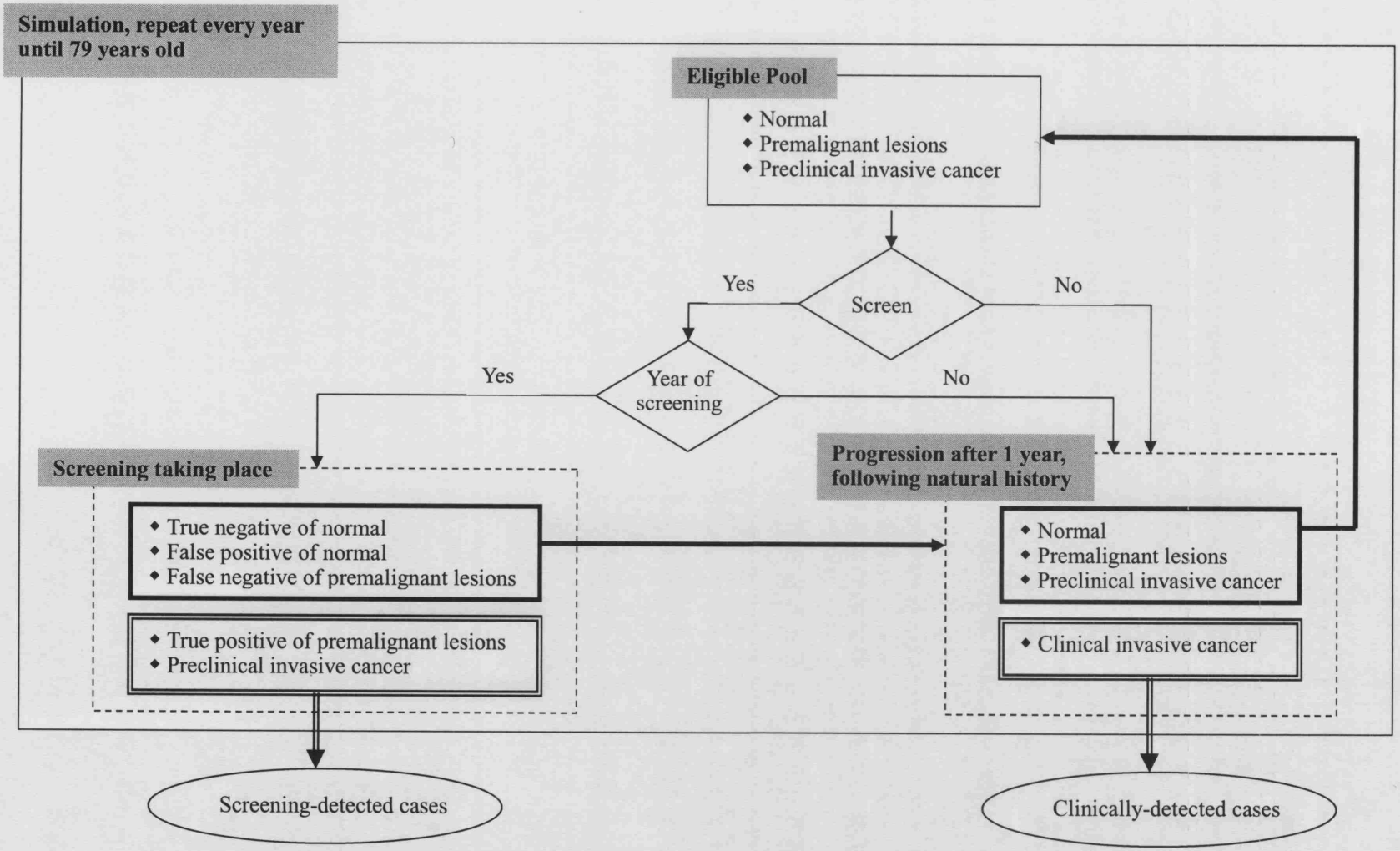


Figure 6.1 The flowchart for computer simulation

6.4 Results

6.4.1 Descriptive results

Table 6.2 shows the distribution of screening findings at the prevalence screen. The proportion with premalignant findings increased with age, from around 1% for women aged under 30 to 4.4% for women aged 80 and over. The detection proportion of invasive cervical cancer had a more radical increase with age, from 0.1 per 1,000 for women aged under 30 to 14 per 1,000 for women aged over 80. When taking age younger than 15 and older than 99 as missing values, the mean (SD) ages of normal, atypical change, low-grade squamous intraepithelial lesion (LSIL), high-grade squamous intraepithelial lesion (HSIL), cervical carcinoma and others at prevalence screen were 41.63 (12.27), 46.13 (13.50), 41.18 (12.53), 48.18 (13.81), 58.91 (13,66), and 48.63 (15.29), respectively.

Table 6.2 Screening findings at prevalence screen by age group

Screening finding	Normal	ASCUS	LSIL	HSIL	Preclinical invasive cancer	Total
15-19	14,519	64	108	15	-	14,706
	98.73%	*0.44%*	*0.73%*	*0.10%*	*0.00%*	
20-29	330,616	1,308	1,686	638	49	334,297
	98.90%	*0.39%*	*0.50%*	*0.19%*	*0.01%*	
30-39	1,009,474	4,972	4,071	3,629	657	1,022,803
	98.70%	*0.49%*	*0.40%*	*0.35%*	*0.06%*	
40-49	721,890	4,881	3,201	3,864	1,199	735,035
	98.21%	0.66%	0.44%	0.53%	0.16%	
50-59	331,013	3,071	1,282	1,875	976	338,217
	97.87%	*0.91%*	*0.38%*	*0.55%*	*0.29%*	
60-69	189,540	2,394	815	1,712	1,154	195,615
	96.89%	*1.22%*	*0.42%*	*0.88%*	*0.59%*	
70-79	69,995	1,081	285	865	612	72,838
	96.10%	*1.48%*	*0.39%*	*1.19%*	*0.84%*	
80-99	10,017	199	65	204	153	10,638
	94.16%	*1.87%*	*0.61%*	*1.92%*	*1.44%*	
Total	2,677,064	17,970	11,513	12,802	4,800	2,724,149
	98.27%	*0.66%*	*0.42%*	*0.47%*	*0.18%*	

* ASCUS: atypical squamous cell of undetermined significance; LSIL: low-grade squamous intraepithelial lesion, encompassing cytological evidence of human papillomavirus/mild dysplasia/cercical intraepithelial neoplasia (CIN) 1; HSIL: high-grade squamous intraepithelial lesion, encompassing moderate and severe dysplasia, carcinoma in situ; CIN 2 and CIN 3

Compared to Table 6.2, the proportion of premalignant findings at subsequent screening is slightly lower but still increases with age, from around 1% for women aged under 30 to 3.5% for women aged 80 and over. The proportion of screen-detected invasive cancer increases with age even more rapidly. (Table 6.3)

Table 6.3 Screening findings at subsequent screening by age group

Screening finding	Normal	ASCUS	LSIL	HSIL	Preclinical invasive cancer	Total
15-19	1,151	3	7	-	-	1,161
	99.14%	*0.26%*	*0.60%*	*0.00%*	*0.00%*	
20-29	85,317	361	392	172	7	86,249
	98.92%	*0.42%*	*0.45%*	*0.20%*	*0.01%*	
30-39	662,758	3,020	2,306	1,720	223	670,027
	98.92%	*0.45%*	*0.34%*	*0.26%*	*0.03%*	
40-49	568,349	3,287	2,078	1,804	322	575,840
	98.70%	*0.57%*	*0.36%*	*0.31%*	*0.06%*	
50-59	223,205	1,723	747	690	160	226,525
	98.53%	*0.76%*	*0.33%*	*0.30%*	*0.07%*	
60-69	86,221	831	355	489	100	87,996
	97.98%	*0.94%*	*0.40%*	*0.56%*	*0.11%*	
70-79	23,564	336	111	199	59	24,269
	97.10%	*1.38%*	*0.46%*	*0.82%*	*0.24%*	
80-99	2,028	33	6	35	11	2,113
	95.98%	*1.56%*	*0.28%*	*1.66%*	*0.52%*	
Total	1,652,593	9,594	6,002	5,109	882	1,674,180
	98.71%	*0.57%*	*0.36%*	*0.31%*	*0.05%*	

Table 6.4 shows the previous screening finding for interval cancers. We can see that the proportion of interval cancers whose previous finding was normal was decreased about half from the youngest age group to the eldest. This could arise from a higher progression rate or a poorer sensitivity in the younger groups.

Table 6.4 Previous screening findings for interval cancers by age group

Screening finding	Normal	ASCUS	LSIL	HSIL	Total
15-19	18	2	-	1	21
	85.71%	*9.52%*	*0.00%*	*4.76%*	
20-29	263	20	11	17	311
	84.57%	*6.43%*	*3.54%*	*5.47%*	
30-39	468	60	18	45	591
	79.19%	*10.15%*	*3.05%*	*7.61%*	
40-49	468	60	18	45	591
	79.19%	*10.15%*	*3.05%*	*7.61%*	
50-59	250	21	2	14	287
	87.11%	*7.32%*	*0.70%*	*4.88%*	
60-69	182	41	5	26	254
	71.65%	*16.14%*	*1.97%*	*10.24%*	
70-79	95	36	4	17	152
	62.50%	*23.68%*	*2.63%*	*11.18%*	
80-99	9	5	1	6	21
	42.86%	*23.81%*	*4.76%*	*28.57%*	
Total	1,285	185	41	126	1,637
	78.50%	*11.30%*	*2.50%*	*7.70%*	

6.4.2 Basic model

Table 6.5 shows the estimated results from the 4-state Markov chain model prior to taking measurement error and heterogeneity of susceptibility of malignant transformation into account. The incidence of premalignant lesions at the subsequent screens increased by age, around 1.4 per 100 women-years for women aged 30-39 to around 2.3 per 100 women-years for women aged over 80. The incidence of premalignant lesions at first screening was non-significantly lower than that at subsequent screening for women aged 30-39 and 70-99, but higher for all other ages and significantly higher for women aged 40-69. The regression rates were more than 1 per women-year for women aged 30-69. Slower regression rates were seen in the older women. The annual transition rates from premalignant lesions detected at incidence screens to preclinical invasive lesions increased with age, from 0.0383 for women aged 30-39 to 0.1452 for women aged 80-99. Taking the inverse of the annual transition rates, we estimated the mean sojourn time in the premalignant state as around 26 years for women aged 30-39 and 7 years for women aged 80-99. Lesions detected at prevalence screen generally had an approximately 50% lower rate of progression. The annual transition rates from preclinical to clinical invasive cancer ranged from 0.1903 to 0.2421, except for a lower estimate of 0.1037 for women aged 70-79. This gives an estimated sojourn time from 4 to 5 years, except for the estimated 9.6 years for women aged 70-79. The proportion of successfully treated premalignancies tended to be lower in women aged over 60.

The results of goodness-of-fit testing are shown in table 6.6. The fit is not good when age groups are combined nor for women aged 30-39, 40-49, 50-59, but is satisfactory for women aged 50-59, 60-69, 70-79, and 80-99. The main discrepancy between observed and fitted comes from the mode of screening-detected invasive cancer, whether with previous normal or premalignant finding, and from the poor fit of interval cancers, as can be seen from the chi-squares at the bottom of table 6.6, when the interval cancers are excluded.

It could be argued that the majority of cases with HSIL or worse are treated by excision and the natural history is not observable thereafter. The same model was therefore applied to data with records truncated after HSIL or more severe lesions were diagnosed. Table 6.7 shows similar results to the inclusive analysis, except for lower estimates of the proportion of successfully treated cases. Similar results on model fit can be seen in table 6.8.

Table 6.5 Estimated results of transitions in cervical cancer screening, Taiwan, 1995-1998

	≥15	≥30	30-39	40-49	50-59	60-69	70-79	80-99
Normal → Premalignant								
Later screening	0.0170	0.0171	0.0142	0.0181	0.0233	0.0250	0.0234	0.0229
	(0.0164-0.0177)	(0.0165-0.0178)	(0.0134-0.0151)	(0.0169-0.0193)	(0.0206-0.0259)	(0.0217-0.0282)	(0.0189-0.0279)	(0.0093-0.0364)
First vs later	1.13	1.15	0.96	1.15	1.19	1.20	0.81	0.99
	(1.10-1.16)	(1.11-1.18)	(0.91-1.01)	(1.09-1.21)	(1.11-1.28)	(1.08-1.34)	(0.59-1.12)	(0.30-3.25)
Premalignant → Normal	1.1948	1.1602	1.0762	1.2309	1.4267	1.1135	0.5716	0.4156
	(1.1234-1.2662)	(1.0869-1.2334)	(0.9636-1.1887)	(1.1005-1.3613)	(1.1910-1.6225)	(0.8831-1.3439)	(0.3001-0.8432)	(-0.3300-1.1611)
Premalignant → Pre-clinical								
Later screening	0.0475	0.0502	0.0383	0.0498	0.0604	0.0664	0.1094	0.1452
	(0.0444-0.0506)	(0.0470-0.0534)	(0.0340-0.0426)	(0.0444-0.0553)	(0.0514-0.0694)	(0.0537-0.0791)	(0.0779-0.1408)	(0.0304-0.2600)
First vs later	0.50	0.59	0.30	0.49	0.70	0.67	0.26	0.51
	(0.40-0.63)	(0.50-0.70)	(0.22-0.41)	(0.37-0.65)	(0.49-1.01)	(0.42-1.07)	(0.12-0.54)	(0.09-2.88)
Pre-clinical → clinical	0.2103	0.2400	0.2239	0.2421	0.2713	0.1903	0.1037	0.2256
	(0.1706-0.2500)	(0.2070-0.2731)	(0.1684-0.2794)	(0.1873-0.2969)	(0.1952-0.3474)	(0.1224-0.2582)	(0.0455-0.1620)	(-0.0187-0.4699)
Pr(Treated)	63%	64%	68%	65%	61%	51%	60%	51%
	(61%-65%)	(62%-66%)	(65%-72%)	(61%-69%)	(54%-68%)	(42%-60%)	(50%-70%)	(21%-81%)

Table 6.6 The comparison of the observed and predicted cases based on 4-state Markov process models by different age groups, Taiwan, 1995-1998

	≥ 15		≥ 30		30-39		40-49	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
First screening								
Normal	2677084	2677083.97	2331947	2331946.96	1009497	1009497.00	721899	721899.03
Premalignancy	42288	42288.01	38459	38459.02	12686	12686.00	11943	11942.97
Pre-clinical	4801	4801.03	4752	4752.03	657	657.00	1199	1199.00
Later screening								
Normal→Normal	1626861	1626643.38	1482182	1481982.67	714705	714644.78	495983	495920.11
Normal→ Premalignancy	16725	16914.74	15501	15682.90	6579	6640.16	5278	5346.28
Normal→ Pre-clinical	693	632.71	679	601.96	226	196.36	240	202.71
Premalignancy →Normal	25734	25864.82	23758	23880.12	9470	9510.23	7994	8033.94
Premalignancy → Premalignancy	3980	3745.27	3721	3498.12	1378	1302.80	1157	1074.70
Premalignancy → Pre-clinical	189	252.70	184	241.12	43	66.91	46	74.73
Interval cancer (IC)								
From Normal	275	194.72	269	219.03	74	47.46	79	65.29
From Premalignancy	77	72.08	75	81.5249	17	16.32	26	22.98
Censoring								
Normal	2636044	2636151.01	2296211	2296309.48	993773	993800.95	710980	711006.28
Premalignancy	32362	32385.63	29359	29377.58	9536	9541.71	8976	8986.36
Model fitting								
Chi-square (6)		72.7885		52.2913		32.9714		28.5748
P-value		5.85E-15		1.20E-10		1.21E-06		9.54E-06
Model fitting excluding IC								
Chi-square (4)		39.3544		40.3688		18.1017		25.2990
P-value		5.89E-08		3.63E-08		1.18E-03		4.38E-05

Table 6.6 (continued)

	50-59		60-69		70-79		80-89	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
First screening								
Normal	331004	331004.03	189537	189537.00	69995	69995.02	10015	10015.00
Premalignancy	6231	6230.97	4909	4909.00	2223	2222.99	467	467.00
Pre-clinical	978	977.99	1153	1153.00	611	610.99	154	154.00
Later screening								
Normal→Normal	183772	183719.92	70099	70081.40	16500	16492.50	1123	1124.06
Normal→ Premalignancy	2262	2307.68	1061	1080.80	298	305.26	23	21.48
Normal→ Pre-clinical	119	106.25	65	55.81	27	25.57	2	2.16
Premalignancy →Normal	3706	3738.79	1961	1972.85	574	580.11	53	52.44
Premalignancy → Premalignancy	515	466.10	464	439.38	182	173.52	25	25.51
Premalignancy → Pre-clinical	36	42.76	33	40.03	22	22.71	4	3.53
Interval cancer (IC)								
From Normal	68	53.97	30	25.69	15	4.23	3	2.56
From Premalignancy	7	18.78	15	14.37	3	3.64	7	3.42
Censoring								
Normal	325778	325805.82	186794	186801.48	68999	69004.40	9887	9885.81
Premalignancy	4650	4644.58	3882	3884.96	1893	1889.88	422	423.73
Model fitting								
Chi-square (6)		19.9806		5.3190		28.2966		4.0295
P-value		0.0005		0.2561		1.09E-05		0.4020
Model fitting excluding IC								
Chi-square (4)		8.9442		4.5683		0.7626		0.2064
P-value		0.0625		0.3345		0.9434		0.9950

Table 6.7 Estimated results of transitions in cervical cancer screening with data truncated after HSIL or severe was diagnosed, Taiwan, 1995-1998

	≥15	≥30	30-39	40-49	50-59	60-69	70-79	80-99
Normal → Premalignant								
Later screening	0.0193	0.0196	0.0163	0.0200	0.0264	0.0298	0.0288	0.0475
	(0.0184-0.0201)	(0.0187-0.0205)	(0.0151-0.0175)	(0.0185-0.0215)	(0.0229-0.0299)	(0.0250-0.0345)	(0.0215-0.0360)	(0.0100-0.0850)
First vs later	1.21	1.27	1.07	1.23	1.29	1.38	1.14	2.03
	(1.18-1.25)	(1.23-1.30)	(1.02-1.12)	(1.17-1.30)	(1.21-1.38)	(1.25-1.51)	(0.90-1.46)	(1.27-3.25)
Premalignant → Normal	1.4774	1.4754	1.3756	1.4672	1.7721	1.5415	1.0047	2.0230
	(1.3836-1.5712)	(1.3776-1.5731)	(1.2297-1.5215)	(1.3050-1.6294)	(1.4624-2.0817)	(1.1976-1.8853)	(0.5517-1.4578)	(0.1498-3.8962)
Premalignant → Pre-clinical								
Later screening	0.0549	0.0486	0.0373	0.0486	0.0593	0.0675	0.1022	0.1178
	(0.0515-0.0583)	(0.0453-0.0518)	(0.0329-0.0416)	(0.0430-0.0541)	(0.0501-0.0686)	(0.0535-0.0815)	(0.0676-0.1369)	(-0.0015-0.2371)
First vs later	0.07	0.59	0.30	0.48	0.72	0.59	0.33	0.44
	(0.06-0.07)	(0.49-0.71)	(0.22-0.41)	(0.36-0.65)	(0.50-1.06)	(0.35-0.98)	(0.14-0.75)	(0.74-3.91)
Pre-clinical → clinical	0.0149	0.2322	0.2166	0.2340	0.2745	0.1701	0.1222	0.1603
	(0.0130-0.0167)	(0.1978-0.2665)	(0.1609-0.2723)	(0.1783-0.2897)	(0.1952-0.3539)	(0.1036-0.2366)	(0.0473-0.1971)	(-0.0611-0.3818)
Pr(Treated)	57%	56%	61%	59%	52%	38%	60%	0%
	(53%-60%)	(53%-60%)	(56%-66%)	(54%-65%)	(40%-63%)	(21%-54%)	(44%-75%)	Lower Bound

Table 6.8 The comparison of the observed and predicted cases based on 4-state Markov process models by different age groups with data truncated after HSIL or severe was diagnosed, Taiwan, 1995-1998

	≥ 15		≥ 30		30-39		40-49	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
First screening								
Normal	2677084	2677083.17	2331947	2331946.99	1009497	1009497.00	721899	721899.00
Premalignancy	42288	42287.80	38459	38459.01	12686	12686.00	11943	11943.00
Pre-clinical	4801	4802.03	4752	4752.00	657	657.00	1199	1199.00
Later screening								
Normal→Normal	1620841	1620476.93	1476479	1476274.72	712321	712260.50	493995	493931.60
Normal→ Premalignancy	16525	16728.02	15309	15501.96	6517	6579.75	5221	5288.88
Normal→ Pre-clinical	688	840.82	675	594.83	226	195.71	236	200.69
Premalignancy →Normal	17779	17879.81	16242	16367.52	6574	6612.18	5444	5485.17
Premalignancy → Premalignancy	2471	2234.43	2292	2073.59	889	816.86	740	662.21
Premalignancy → Pre-clinical	113	245.83	110	170.09	25	50.13	27	53.36
Interval cancer (IC)								
From Normal	273	11.49	267	220.20	74	47.35	79	64.37
From Premalignancy	43	4.74	41	92.17	11	18.00	15	25.58
Censoring								
Normal	2628611	2628868.06	2289197	2289301.50	991054	991083.75	708598	708626.63
Premalignancy	23423	23461.92	20869	20868.14	7146	7147.78	6355	6358.29
Model fitting								
Chi-square (6)		6388.5359		96.7911		42.2035		37.2647
P-value		0.00E+00		4.74E-20		1.51E-08		1.59E-07
Model fitting excluding IC								
Chi-square (4)		127.7985		58.4365		24.4819		29.5637
P-value		1.15E-26		6.18E-12		0.0001		6.00E-06

Table 6.8 (continued)

	50-59		60-69		70-79		80-89	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
First screening								
Normal	331004	331004.00	189537	189537.00	69995	69995.00	10015	10015.00
Premalignancy	6231	6231.00	4909	4909.00	2223	2223.00	467	467.00
Pre-clinical	978	978.00	1153	1153.00	611	611.00	154	154.00
Later screening								
Normal→Normal	182952	182902.30	69692	69672.26	16404	16396.92	1115	1114.92
Normal→ Premalignancy	2219	2263.15	1039	1058.70	291	298.98	22	21.76
Normal→ Pre-clinical	119	105.11	65	58.40	27	24.20	2	2.10
Premalignancy →Normal	2600	2630.28	1240	1254.29	352	357.68	32	32.37
Premalignancy → Premalignancy	329	283.37	246	224.06	78	69.48	10	9.68
Premalignancy → Pre-clinical	24	31.61	24	27.63	8	9.93	2	1.76
Interval cancer (IC)								
From Normal	67	57.73	29	24.21	15	6.39	3	2.78
From Premalignancy	4	23.13	8	14.57	1	3.58	2	1.94
Censoring								
Normal	324761	324787.86	186122	186129.97	68794	68798.40	9868	9868.19
Premalignancy	3380	3373.99	2583	2582.69	1173	1170.43	232	232.06
Model fitting								
Chi-square (6)		29.5618		7.8168		15.5166		0.0742
P-value		6.01E-06		0.0985		0.0037		0.9993
Model fitting excluding IC								
Chi-square (4)		12.2515		3.9065		2.0560		0.0550
P-value		0.0156		0.4188		0.7255		0.9996

6.4.3 Measurement error and Mover-Stayer model

Table 6.9 shows the estimated results of the model taking measurement error and heterogeneity of progression capability into account. In this model, only data from subsequent screens were included. The incidence of premalignant lesions increased by age from 2 per 1000 woman-years for women aged 30-49 to 6 per 1000 woman-years for women in their sixties. The regression rates also increased with age among women aged 30-69. The malignant transition rates from premalignant lesions among those with susceptibility (movers) were estimated as 0.77 in women aged 30-39, and around 1.2 for women aged 40-69. The annual transition rate from preclinical to clinical invasive cancer increased by age from 30-59, and decreased after 60. The estimated proportion of movers increased with age from 6% to 11%. The estimated results show good performance of screening in terms of measurement error with 100% sensitivity and above 98% negative predictive value in all age groups. The estimates, however, were quite unstable for the age group 70-79 in this model, being very dependent on starting values in the estimation.

Table 6.10 shows the result of goodness-of-fit testing in this model. The results show a good fit in all age groups, except for age group 40-49, with a Pearson χ^2 of 9.93 with 2 degrees of freedom (p-value 0.007). This is mainly due to underestimation of interval cancers after a previous premalignant screening finding.

Table 6.9 Estimated transition rates and other parameters in cervical cancer screening with measurement error and heterogeneity in the model, Taiwan, 1995-1998

Parameters	30-39	40-49	50-59	60-69	70-79
Normal→Premalignant	0.0023	0.0022	0.0047	0.0059	0.0021
	(0.000*-0.0101)	(0.000*-0.0073)	(0.000*-0.0276)	(0.000*-0.0648)	(0.000*-0.0233)
Premalignant→Normal	0.2035	0.2359	0.3790	0.4174	0.0661
	(0.0847-0.3223)	(0.1122-0.3595	(0.1522-0.6058)	(0.0956-0.7393)	(0.000*-0.3780)
Premalignant→Preclinical	0.7714	1.1084	1.2467	1.1712	2.0926
	(0.1312-1.4116)	(0.3385-1.8782	(0.2218-2.2715)	(0.000*-2.5481)	(0.000*-4.5076)
Preclinical→Clinical	0.1778	0.1862	0.2108	0.1582	0.0891
	(0.1346-0.2209)	(0.1434-0.2290	(0.1517-0.2698)	(0.1009-0.2155)	(0.0402-0.1381)
Pr (treated)	84%	84%	82%	70%	72%
	(31%-100%*)	(44%-100%*)	(0%-100%*)	(0%-100%*)	(0%-100%*)
Sensitivity	100%	100%	100%	100%	100%
	(---)	(---)	(---)	(---)	(---)
NPV	99%	99%	99%	98%	98%
	(96%-100%*)	(96%-100%*)	(92%-100%*)	(82%-100%*)	(80%-100%*)
Pr (mover)	6%	6%	7%	8%	11%
	(0%*-26%)	(0%*-21%)	(0%*-42%)	(0%*-86%)	(0%*-100%)

* Truncated according to the parameter space

Table 6.10 The comparison of the observed and predicted cases based on 4-state Markov process models with measurement error and heterogeneity considerations by different age groups, Taiwan, 1995-1998

Mode	30-39		40-49		50-59		60-69		70-79	
	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected	Observed	Expected
Normal → Normal	714705	714675.88	495983	495955.34	183772	183743.92	70099	70093.32	16500	16495.46
Normal → Premalignant	6579	6574.23	5278	5269.81	2262	2259.27	1061	1054.09	298	296.52
Normal → Preclinical	226	225.55	240	236.85	119	126.30	65	68.18	27	30.52
Premalignant → Normal	9470	9460.20	7994	7978.33	3706	3705.76	1961	1955.08	574	574.75
Premalignant → Premalignant	1378	1380.19	1157	1162.10	515	516.68	464	468.98	182	183.63
Premalignant → Preclinical	43	44.81	46	49.60	36	29.52	33	30.24	22	18.35
Interval cancer										
From normal	74	79.29	79	90.42	68	63.61	30	33.9827	15	12.89
From premalignant	17	11.86	26	14.99	7	11.69	15	11.20	3	5.30
Censored										
From normal	993773	993802.30	710980	711007.84	325778	325806.10	186794	186799.52	68999	69003.64
From premalignant	9536	9546.98	8976	8994.02	4650	4650.34	3882	3889.51	1893	1891.95
Chi-squared		2.69		9.93		4.04		2.29		2.50
P-value		0.26		0.0070		0.1323		0.3189		0.29

6.4.4 Computer simulation

Table 6.11 shows the simulated results of different screening regimes using the estimated parameters of the model with heterogeneity (Table 6.9) by screening interval and starting age under the assumption of no previous screening. Four situations were tabulated: sensitivity of 95% and 90%, and proportions of successfully treated premalignancies of 100% and 85% (i.e. assuming a more aggressive treatment policy than in the programme). In table 6.11 it is shown that the reduction in invasive cancers (screening-detected or clinically detected), compared with no screening at all, decreased with a longer screening interval. When screening starts at age 30 years, the reduction decreases from 72% with annual screening to 20% with ten-yearly. When screening starts at 60 years, the change in the reduction becomes less radical, from 53% with annual screening to 21% with 10-yearly screening. Under the same screening interval, the reduction is smaller when screening starts later in life, but screening is more costly when screens start at from younger ages, due to the longer period and therefore larger number of screens. In order to take both cost and effectiveness into account, a ratio of smears required per invasive cancer prevented, S/I, was used. With the same starting age, a longer screening interval leads to smaller S/I. There is clearly a trade-off between the effectiveness in terms of absolute rates of invasive cases detected and marginal cost in terms of number of smears required per invasive case prevented. It could be argued that 3-yearly screening confers a substantial benefit of the order of 40-50% of invasive cancer prevented, with an acceptable marginal cost. Annual screening confers a clear additional absolute benefit, but at a substantially greater absolute and marginal cost. The smaller the sensitivity or proportion successfully treated, the larger the figure of S/I.

Table 6.12 shows the simulated results of different screening regimes by screening interval and starting age under the assumption of perfect screening beforehand. The results indicate

greater absolute benefits but larger S/I than in Table 6.11. The qualitative implications for absolute benefits and marginal costs, however, are similar.

Table 6.11-1 The result of computer simulation assuming no previous screening (Sen=95%, Spe=98%, Pr(Tx)=100%) by screening age involved and screening interval

Age and interval / Parameters	30-79					40-79					50-79					60-79				
	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*
N. False positive	993.68	335.81	196.94	98.24	0.00	794.74	276.41	157.55	78.64	0.00	595.27	197.13	117.93	58.88	0.00	396.73	138.04	78.64	39.26	0.00
N. Premalignant detected	148.19	107.69	80.75	47.55	0.00	126.05	91.46	66.23	37.73	0.00	108.34	75.77	56.57	32.54	0.00	72.04	53.06	39.54	23.58	0.00
N. Invasive cancer	6.68	12.86	16.25	18.91	23.64	6.23	12.55	14.72	16.71	20.06	6.37	11.42	13.67	14.98	17.46	5.83	9.14	10.70	10.99	12.36
N. Preclinical invasive ca.	6.21	10.78	12.15	9.88	0.00	5.78	10.60	11.03	8.67	0.00	5.83	9.58	10.35	7.99	0.00	5.34	7.90	8.53	6.43	0.00
N. Clinical invasive ca.	0.47	2.08	4.10	9.03	23.64	0.44	1.95	3.69	8.04	20.06	0.54	1.84	3.32	6.99	17.46	0.49	1.24	2.17	4.56	12.36
No. of screens	49847	16916	9946	4973	0	39876	13929	7959	3982	0	29884	9947	5968	2987	0	19918	6966	3983	1995	0
Marginal cost^	2938	1569	1345	1050		2882	1854	1490	1188		2696	1648	1575	1207		3049	2163	2398	1455	

* C: control group—no screens

^ Number of pap smears required to prevent one invasive cervical cancer compared to not screening at all

Table 6.11-2 The result of computer simulation assuming no previous screening (Sen=90%, Spe=98%, Pr(Tx)=100%) by screening age involved and screening interval

Age and interval / Parameters	30-79					40-79					50-79					60-79				
	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*
N. False positive	993.39	335.68	196.88	98.23	0.00	794.51	276.31	157.50	78.64	0.00	595.09	197.06	117.90	58.87	0.00	396.62	137.99	78.62	39.26	0.00
N. Premalignant detected	145.63	103.94	77.24	45.11	0.00	123.76	88.17	63.27	35.77	0.00	106.26	72.92	54.00	30.84	0.00	70.69	51.15	37.77	22.34	0.00
N. Invasive cancer	7.09	13.52	16.84	19.27	23.64	6.61	13.15	15.24	17.00	20.06	6.71	11.91	14.12	15.23	17.46	6.08	9.55	11.05	11.18	12.36
N. Preclinical invasive ca.	6.59	11.29	12.52	9.95	0.00	6.13	11.06	11.35	8.73	0.00	6.14	9.96	10.64	8.03	0.00	5.57	8.23	8.76	6.46	0.00
N. Clinical invasive ca.	0.51	2.23	4.32	9.31	23.64	0.48	2.08	3.89	8.28	20.06	0.58	1.96	3.49	7.21	17.46	0.51	1.32	2.29	4.72	12.36
No. of screens	49839	16912	9944	4973	0	39870	13926	7958	3981	0	29879	9945	5967	2987	0	19916	6966	3983	1995	0
Marginal cost^	3011	1671	1462	1136		2963	2013	1649	1302		2781	1794	1789	1343		3170	2475	3043	1691	

* C: control group—no screens

^ Number of pap smears required to prevent one invasive cervical cancer compared to not screening at all

Table 6.11-3 The result of computer simulation assuming no previous screening (Sen=95%, Spe=98%, Pr(Tx)=85%) by screening age involved and screening interval

Age and interval / Parameters	30-79					40-79					50-79					60-79				
	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*
N. False positive	992.81	335.43	196.77	98.20	0.00	794.04	276.11	157.43	78.62	0.00	594.71	196.93	117.85	58.86	0.00	396.40	137.90	78.59	39.26	0.00
N. Premalignant detected	165.13	113.69	83.01	47.75	0.00	140.06	96.22	67.84	37.81	0.00	120.02	79.32	57.80	32.57	0.00	79.92	55.82	40.50	23.60	0.00
N. Invasive cancer	7.95	14.82	17.96	19.94	23.64	7.39	14.32	16.21	17.55	20.06	7.42	12.88	14.98	15.70	17.46	6.60	10.35	11.72	11.54	12.36
N. Preclinical invasive ca.	7.37	12.29	13.21	10.09	0.00	6.84	11.98	11.96	8.82	0.00	6.78	10.70	11.17	8.09	0.00	6.05	8.88	9.20	6.51	0.00
N. Clinical invasive ca.	0.58	2.53	4.75	9.85	23.64	0.55	2.34	4.26	8.73	20.06	0.64	2.18	3.81	7.60	17.46	0.55	1.47	2.52	5.03	12.36
No. of screens	49822	16905	9941	4972	0	39857	13920	7956	3981	0	29869	9942	5966	2987	0	19911	6964	3982	1995	0
Marginal cost^	3175	1915	1750	1341		3145	2422	2066	1583		2975	2172	2405	1695		3453	3455	6174	2419	

* C: control group—no screens

^ Number of pap smears required to prevent one invasive cervical cancer compared to not screening at all

Table 6.11-4 The result of computer simulation assuming no previous screening (Sen=90%, Spe=98%, Pr(Tx)=85%) by screening age involved and screening interval

Age and interval / Parameters	30-79					40-79					50-79					60-79				
	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*
N. False positive	992.51	335.31	196.71	98.19	0.00	793.80	276.02	157.39	78.61	0.00	594.52	196.86	117.82	58.86	0.00	396.28	137.86	78.57	39.26	0.00
N. Premalignant detected	161.98	109.53	79.30	45.29	0.00	137.25	92.60	64.73	35.84	0.00	117.50	76.22	55.11	30.86	0.00	78.28	53.73	38.64	22.36	0.00
N. Invasive cancer	8.39	15.44	18.49	20.25	23.64	7.79	14.88	16.67	17.80	20.06	7.78	13.35	15.37	15.91	17.46	6.86	10.74	12.03	11.70	12.36
N. Preclinical invasive ca.	7.77	12.77	13.54	10.15	0.00	7.20	12.42	12.24	8.86	0.00	7.11	11.06	11.42	8.13	0.00	6.29	9.19	9.41	6.53	0.00
N. Clinical invasive ca.	0.62	2.67	4.96	10.09	23.64	0.59	2.46	4.43	8.94	20.06	0.67	2.28	3.96	7.78	17.46	0.57	1.54	2.62	5.17	12.36
No. of screens	49814	16902	9939	4971	0	39850	13918	7955	3981	0	29864	9940	5965	2986	0	19908	6963	3982	1995	0
Marginal cost^	3266	2062	1929	1463		3247	2687	2344	1759		3086	2418	2865	1929		3619	4282	11891	3014	

* C: control group—no screens

^ Number of pap smears required to prevent one invasive cervical cancer compared to not screening at all

Table 6.12-1 The result of computer simulation assuming intensive screening previously (Sen=95%, Spe=98%, Pr(Tx)=100%) by screening age involved and screening interval

Age and interval / Parameters	30-79					40-79					50-79					60-79				
	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*
N. False positive	994.79	336.31	197.30	98.51	0.00	795.62	276.81	157.83	78.85	0.00	596.40	197.60	118.28	59.12	0.00	397.90	138.55	79.00	39.53	0.00
N. Premalignant detected	139.98	99.57	72.69	39.54	0.00	119.70	85.18	59.98	31.52	0.00	101.13	68.65	49.48	25.44	0.00	64.14	45.27	31.78	15.81	0.00
N. Invasive cancer	5.74	11.92	15.29	17.94	21.55	5.30	11.61	13.78	15.76	18.16	4.75	9.79	12.03	13.35	14.96	3.33	6.63	8.19	8.47	9.21
N. Preclinical invasive ca.	5.40	9.97	11.34	9.07	0.00	4.99	9.80	10.23	7.88	0.00	4.48	8.23	9.01	6.65	0.00	3.17	5.72	6.36	4.27	0.00
N. Clinical invasive ca.	0.34	1.95	3.95	8.87	21.55	0.31	1.81	3.54	7.87	18.16	0.27	1.56	3.03	6.70	14.96	0.16	0.90	1.83	4.20	9.21
No. of screens	49893	16931	9954	4977	0	39912	13941	7966	3985	0	29931	9962	5976	2990	0	19966	6982	3991	1998	0
Marginal cost^	3156	1758	1591	1379		3103	2128	1815	1657		2932	1926	2041	1850		3394	2704	3894	2706	

* C: control group—no screens

^ Number of pap smears required to prevent one invasive cervical cancer compared to not screening at all

Table 6.12-2 The result of computer simulation assuming intensive screening previously (Sen=90%, Spe=98%, Pr(Tx)=100%) by screening age involved and screening interval

Age and interval / Parameters	30-79					40-79					50-79					60-79				
	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*
N. False positive	994.53	336.19	197.25	98.50	0.00	795.40	276.72	157.79	78.85	0.00	596.23	197.54	118.25	59.12	0.00	397.81	138.51	78.99	39.52	0.00
N. Premalignant detected	137.52	96.05	69.49	37.50	0.00	117.49	82.09	57.28	29.87	0.00	99.20	66.09	47.24	24.11	0.00	62.96	43.69	30.39	14.98	0.00
N. Invasive cancer	6.14	12.54	15.84	18.25	21.55	5.67	12.18	14.25	16.01	18.16	5.08	10.25	12.45	13.55	14.96	3.56	7.00	8.49	8.61	9.21
N. Preclinical invasive ca.	5.77	10.45	11.68	9.13	0.00	5.33	10.24	10.53	7.92	0.00	4.78	8.58	9.27	6.68	0.00	3.38	6.02	6.57	4.28	0.00
N. Clinical invasive ca.	0.37	2.09	4.16	9.12	21.55	0.34	1.93	3.72	8.08	18.16	0.30	1.67	3.18	6.87	14.96	0.18	0.97	1.93	4.33	9.21
No. of screens	49885	16928	9953	4977	0	39906	13938	7965	3984	0	29927	9960	5976	2990	0	19964	6981	3990	1998	0
Marginal cost^	3238	1879	1744	1507		3193	2329	2037	1845		3027	2113	2375	2117		3531	3151	5547	3340	

* C: control group—no screens

^ Number of pap smears required to prevent one invasive cervical cancer compared to not screening at all

Table 6.12-3 The result of computer simulation assuming intensive screening previously (Sen=95%, Spe=98%, Pr(Tx)=85%) by screening age involved and screening interval

Age and interval / Parameters	30-79					40-79					50-79					60-79				
	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*
N. False positive	993.98	335.97	197.16	98.48	0.00	794.96	276.54	157.73	78.84	0.00	595.88	197.43	118.21	59.12	0.00	397.61	138.44	78.96	39.52	0.00
N. Premalignant detected	155.84	104.96	74.59	39.65	0.00	132.91	89.53	61.37	31.56	0.00	112.07	71.92	50.59	25.46	0.00	71.24	47.77	32.64	15.82	0.00
N. Invasive cancer	6.97	13.77	16.88	18.81	21.55	6.42	13.29	15.15	16.46	18.16	5.74	11.15	13.22	13.93	14.96	4.03	7.72	9.07	8.87	9.21
N. Preclinical invasive ca.	6.52	11.40	12.32	9.23	0.00	6.01	11.11	11.09	7.99	0.00	5.38	9.28	9.76	6.73	0.00	3.82	6.62	6.95	4.30	0.00
N. Clinical invasive ca.	0.45	2.37	4.56	9.58	21.55	0.41	2.17	4.06	8.47	18.16	0.36	1.86	3.46	7.20	14.96	0.21	1.10	2.11	4.58	9.21
No. of screens	49870	16922	9950	4976	0	39895	13934	7964	3984	0	29918	9957	5975	2990	0	19959	6979	3990	1997	0
Marginal cost^	3421	2176	2133	1819		3397	2857	2645	2337		3245	2609	3431	2887		3854	4676	27766	5901	

* C: control group—no screens

^ Number of pap smears required to prevent one invasive cervical cancer compared to not screening at all

Table 6.12-4 The result of computer simulation assuming intensive screening previously (Sen=90%, Spe=98%, Pr(Tx)=85%) by screening age involved and screening interval

Age and interval / Parameters	30-79					40-79					50-79					60-79				
	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*	1	3	5	10	C*
N. False positive	993.71	335.86	197.11	98.47	0.00	794.74	276.46	157.70	78.83	0.00	595.71	197.37	118.19	59.12	0.00	397.51	138.40	78.95	39.52	0.00
N. Premalignant detected	152.82	101.07	71.22	37.59	0.00	130.21	86.14	58.54	29.91	0.00	109.72	69.12	48.25	24.13	0.00	69.81	46.02	31.17	14.99	0.00
N. Invasive cancer	7.39	14.37	17.37	19.07	21.55	6.81	13.83	15.57	16.67	18.16	6.08	11.58	13.58	14.10	14.96	4.28	8.07	9.34	8.99	9.21
N. Preclinical invasive ca.	6.91	11.87	12.62	9.28	0.00	6.36	11.54	11.36	8.02	0.00	5.70	9.62	9.99	6.75	0.00	4.04	6.90	7.13	4.30	0.00
N. Clinical invasive ca.	0.49	2.50	4.75	9.80	21.55	0.44	2.29	4.22	8.65	18.16	0.39	1.96	3.59	7.35	14.96	0.23	1.17	2.20	4.69	9.21
No. of screens	49863	16919	9949	4976	0	39889	13931	7963	3984	0	29913	9956	5974	2990	0	19957	6979	3990	1997	0
Marginal cost^	3523	2356	2381	2011		3512	3211	3074	2664		3369	2944	4331	3467		4043	6124	<0	9113	

* C: control group—no screens

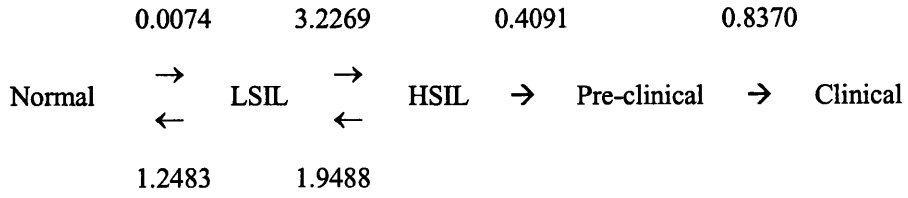
^ Number of pap smears required to prevent one invasive cervical cancer compared to not screening at all

6.5 Discussion

In this chapter, we started from a basic model of disease progression of cervical cancer applied to the population screening programme in Taiwan. The results suggested that the low malignant transition rates estimated were due to ignoring heterogeneity of susceptibility. From the estimated results numbers of predicted screen-detected carcinoma cases at incidence screens from previous disease-free findings were underestimated and numbers from previous premalignant conditions were overestimated. It was suspected that overestimated cases resulted from false-negative cases, i.e. the sensitivity or the negative predictive value cannot be 100%. Interval cancer rates after a previous normal finding were similarly underestimated. Therefore, mover-stayer models incorporating measurement error and heterogeneity of potential transformation from the premalignant lesions into the model were developed. The latter showed a significant improvement in model fit.

In the final mover-stayer model with measurement error, only data form the incidence (second or subsequent) screens were used. Prior to this, we tried a mover-stayer model taking into account measurement error and length bias simultaneously, with prevalence and incidence data in the same model. The estimates were very sensitive to the initial values of sensitivity and proportion of premalignant cases successfully treated. Also, different sets of estimates gave very similar values of the log-likelihood. This suggested identification problems due to overparameterisation.

One point which is illustrated by the results in this chapter is that the most complex model is not necessary the best fitting. A more complex disease process with LSIL and HSIL as different disease state has shown worse fit than our final model. The following is an example for women aged 50-59:



Estimate of the proportion of successfully treated or misclassified LSIL cases: 84%

Estimate of the proportion of successfully treated or misclassified HSIL cases: 85%

Table 6.13 The comparison of observed and predicted cases based on a 5-state Markov process model for women aged 50-59, Taiwan, 1995-1998

	Observed	Expected
First screening		
Normal	332746	332750.79
LSIL	1299	1359.72
HSIL	1912	1862.60
Preclinical	929	912.89
Later screening		
Normal→ Normal	185146	185066.55
Normal→ LSIL	523	487.05
Normal→ HSIL	446	515.83
Normal→ Preclinical	115	105.21
LSIL → Normal	887	885.87
LSIL → LSIL	64	27.75
LSIL → HSIL	33	49.92
LSIL → Preclinical	2	14.10
HSIL → Normal	1083	1080.66
HSIL → LSIL	37	35.33
HSIL → HSIL	93	71.05
HSIL → Preclinical	12	23.99
Interval cancer (IC)		
From Normal	75	123.52
From LSIL	1	11.25
From HSIL	3	18.43
Model fitting		
Chi-square	135.01	
P-value	<0.0000	

Similar results were observed in a 5-state mover-stayer model, with a chi-square of 181.76 for goodness of fit.

The fact that the annual premalignant incidence and regression rates were about 5 times larger in the basic model than in the heterogeneity model can be explained by the failure to take heterogeneity into account. All premalignant lesions were assumed to be capable of progression, so that higher regression rates were estimated, which in turn led to higher estimated incidence rates of premalignant lesions.

The proportion of movers who have the potential to follow the entire natural history of cervical cancer from normal to invasive carcinoma was estimated as between 6% and 11% in our study. This estimate was lower than the finding of 24% progression in women aged 25-50 by Bos et al.[149] Van Oortmarssen and Habbema concluded that 84% of new lesions regress spontaneously for women aged under 34, and 40% for women over age 34. Their study also found a larger proportion of progressive lesions than ours. Raffle and colleagues' study[139] showed a closer result to ours. They found that at least 80% of high grade dyskaryosis and of high grade dysplasia would not progress to cancer. The discrepancies from our study might be explained by the fact that our study was based on an annual screening data, which would therefore diagnose more non-progressive premalignant lesions due to the frequent screens.

In the screening assessment, we used a ratio of number of smears per invasive cases prevented (S/I) to represent the marginal cost of screening. In the results of the simulations, there is a further benefit of annual screening over 3-yearly but at a considerable increase in both absolute and marginal costs. The simulation results also suggest that for screening with a 3-year interval or longer, screening starting from 30 years old is the optimum option with respect of interval cancer reduction and S/I. Moreover, the choice of interval also depends on societal willingness to attend. Under the policy of annual screening in Taiwan during 1995-1998, there were only 43% attending repeated pap screening in the three years of observation. The average time since last screen, for the women who actually did attend for

repeat screening, was 1.4 years.

Generally speaking, a disease natural history model has to take heterogeneity into account. Very simple homogeneous models estimate sojourn times of 12-20 years. This would mean that 5-10 yearly screening would be adequate. Our chosen model fitted the observed data much better than a homogeneous model, and simulations from our model showed a clear and substantial benefit of 3-yearly screening over 5-yearly. Sasieni et al[150], on the basis of case-control evidence, proposed a 3-year interval for women aged 25-49 in the UK. This recommendation has been implemented.[150;151]

Chapter 7 Discussion and Conclusion

7.1 Generic implications—methodological

There is no general rule about whether a simple or more complex model will give a better fit. For example, in chapter 2, the more complex stochastic model clearly performed better in modelling DCIS progression than the simple deterministic model. Similarly, a heterogeneous model gave a better fit than a homogeneous to the cervical screening data in chapter 6. However, additional complexity beyond the 4-state heterogeneous model for cervical cancer led to a deterioration in fit rather than a further improvement.

The results of chapter 4 showed only a slightly better fit of heterogeneous models over homogeneous, in the analysis of the adenoma-carcinoma process in colorectal screening. However, the better fit was most strongly manifested in the estimation of numbers of colorectal cancers, i.e. the most important event in clinical terms. The heterogeneity models also gave more plausible prediction of further cumulative disease progression.

In view of the results of chapters 2, 4, and 6, therefore, it is clear that there is a potential role for heterogeneous models in cancer screening. The two practical questions which arise are (1) the ability to fit the model and estimate the parameters with precision and consistency; (2) the usefulness of the results in terms of policy decisions.

(1) will depend on the complexity of the model, the methodology, and the data available. (2) depends on the fit of the model and its plausibility based on knowledge of the biological process being modelled. We now consider the two questions in more detail.

7.1.1 Tractability and estimation

As in other fields, the more complex the model, the more difficulty in fitting. This is not a question of simple overparameterisation. In various examples we found either difficulties in convergence, dependence on starting values or non-invertability of the matrix of second derivatives, despite there being theoretically sufficient degrees of freedom in the data. This illustrates the fact that a complex model can be identifiable in principle, but may require more data than one could hope to gather in practice for reliable estimation. This is the case for some of the more complex models attempted for chapter 6 (numerous models were tried but their results are not reported here). Despite having screening data on 2.7 million women, some models gave results which were strongly dependent on starting values.

While a model can be too complex, it can also be too simple, as noted for the cervical screening example. Also, the results in chapter 4 are suggestive of better fit and greater plausibility of the frailty models than the homogeneous models to the colorectal data.

Chapter 4 also raises possibly the most important methodological part of this work, the issue of tractability of the models. Superficially, applying continuous frailty to multistate disease processes (particularly if there are 3 states or more) is a daunting prospect. This is because there is no longer a straightforward matrix approach to deriving finite transition probabilities, so that we rely on complex multidimensional numerical integration. However, the strategy of splitting up the process into two piecewise homogeneous processes, separated by a single transition with frailty, proved effective. In principle, this approach could be extended to frailty in more than one transition, separating more than two piecewise homogeneous processes. We suspect, however, that this would run into computing and estimation problems.

7.1.2 Plausibility of the model and usefulness of the results

To some extent, these are dependent on both the model and the data. A model may be correct in biological terms but the data may be insufficient to give precise estimates or indeed to fully identify the parameters. On the other hand, a model may give a good fit to the data, but yield estimates which we do not believe to be plausible. The casting calcification work in chapter 5 is a good example. In chapter 5, several models gave a good fit to the data but all yielded at least one estimate which was not plausible in clinical or biological terms. This may be partly due to the fact that in a large number of subjects, the crucial variable (presence of casting-type calcifications) was either unobserved or unobservable, but it is more likely to be due to the fact that the biological process is not well understood, and the models do not adequately reflect this process. This illustrates the fact that with sufficient parameters one can obtain a good fit to the data, even if the model is unrealistic. This in turn is a warning against overcomplexity.

7.1.3 Conclusions on methodological strategies

Clearly, a good rule of thumb is that one should start with the simplest models and gradually introduce increasing complexity, assessing the improvement, if any, at each stage. It may be that there is no advantage in moving beyond the simplest model. It is more likely, however, that an improvement in fit will be observed initially with increasing complexity, but at some point the improvement will cease to be observed, or problems of estimation will arise. In this case a "best" model of intermediate complexity will be observed, as in the cervical screening example. The work in chapter 4 developing piecewise homogeneous models separated by a transition incorporating frailty gives an opportunity to introduce continuous frailty to multistate models as part of this process.

7.2 Specific implications—clinical

7.2.1 Overdiagnosis of DCIS in breast screening

In addition to the debate on the role of detecting ductal carcinoma in situ (DCIS) in mammographic screening between 'mainly overdiagnosis' and ' a representation of substantial prevention of invasive breast cancer', our results suggested that there is an element of overdiagnosis and overtreatment of DCIS in mammographic screening programmes. We estimated that 37% and 4% of DCIS cases were non-progressive when diagnosed at a prevalence screen and an incidence screen, respectively. Since DCIS cases are a minority of screen-detected cancers, this means that the great majority of cases diagnosed at screening are progressive, either invasive disease or progressive DCIS. We estimated that for each non-progressive DCIS case diagnosed at a prevalence screen, 19 invasive or progressive DCIS cases are diagnosed. For incidence screens, there are around 140 progressive DCIS or invasive cases diagnosed per single non-progressive case. Although we should not be complacent and we should always strive to minimise unnecessary intervention, this suggests that the overdiagnosis element is modest in comparison with the likely benefit of early diagnosis and treatment of progressive lesions. The finding that there is an element of overdiagnosis also raises the question of treatment of DCIS. Clearly, it would be imprudent not to excise such tumours given the finding that the majority will progress to invasive breast cancer if left untreated. However, there is an urgent need to design the adjuvant treatment to fit the risk the disease poses to the patient. In particular, classification of DCIS cases into those who need postoperative radiotherapy and those who do not would be of great value. Thus our results also pose a challenge to therapy of DCIS as much as to early detection.

7.2.2 Colonoscopy screening

With the incorporation of frailty in a general multi-state model with continuous or dichotomous heterogeneity for the natural history of the adenoma-carcinoma sequence in the

colon and rectum, we found consistent results of three-state models, implying that around 40% of adenomas would progress to invasive cancer in 20 years. In a more complex model with adenoma states classified by polyp size or histological type, a very high early cumulative risk among the progressive or "faster' cases from diminutive or tubular adenomas was seen. This could result from the assumption of only diminutive (or tubular) adenomas having lower or zero susceptibility, so that early events are dominated by those with greater susceptibility.

7.2.3 Cervical cancer

In the investigation of heterogeneity of abnormalities detected in cervical cancer screening, around 6%-11% of premalignancies were estimated to have positive susceptibility of malignant transformation. Our empirical estimates of disease progression rates were from a model which fitted well to screening data on 2.7 million women. The results of computer simulations from these estimates showed a clear and substantial benefit of a 3-yearly screening interval in terms of marginal cost of screening by using numbers of smears required to prevent one invasive cervical cancer. This is borne out by study of invasive cancer by time since last screen (Table 7.1). Clearly, after 3 years since last screen, the incidence of invasive cancer increases dramatically.

Table 7.1 Invasive cervical cancer incidence by time since last screen

Time since last screen (year)	Numbers of Invasive cancer			Person-years	Incidence Rate per 1000		
	Screen-detected	IC	Total		Screen-detected	IC	Total
0.50-0.99	303	176	479	719856.3	0.4209	0.2445	0.6654
1.00-1.99	460	130	590	499890.7	0.9202	0.2601	1.1803
2.00-2.99	94	40	134	85665.8	1.0973	0.4669	1.5642
3.00-4.00	25	6	31	5759.0	4.3410	1.0418	5.3829

Results from homogeneous models suggest a very slow average progression rate of premalignancy and suggest that longer interscreening intervals are acceptable. Assuming that all premaligancies can progress when in fact the majority cannot means that the estimated rate of progression and therefore screening interval policy, will be based mainly on the majority of lesions which do not need screen-detection, rather than on the minority which do.

7.3 Future work

From a methodological point of view, there is no doubt that pure statistical research will continue to develop methods and models for multistate data with frailties. The piecewise homogeneity strategy seems to be a good way forward.

For models to be of any use, there must be data. Our results emphasise the need for rich and high-quality data on disease events in screening programmes. The results on casting-type calcifications illustrate the difficulties of estimation when there are major gaps in the data, notably on a key variable. An immediate challenge for those working in screening programmes is collection of data not only for monitoring the programmes but also for evaluation of hypotheses which may lead to improved screening and disease control in the future.

In clinical terms, there are a number of targets for the future. While recent work by others and in this thesis has indicated optimum screening regimes for pap smear screening, interest in human papillomavirus (HPV) testing as a frontline screening tool is increasing. There will be a need to evaluate screening using HPV testing in the near future.

There remains considerable work to be done in modelling and evaluation of the many

colonoscopy programmes in individuals at high risk of colorectal cancer around the world. Other targets for evaluation in the future will be the evaluation and comparison of colonoscopy, sigmoidoscopy and virtual colonoscopy.

Lung cancer screening with computed tomography is a topic of some interest. Because of fears of overdiagnosis, there will be a need for detailed statistical modelling and analysis to investigate this.

When breast cancer screening was initiated, the clinical and scientific worlds were given the opportunity for the first time to study the natural history of early invasive breast cancer and DCIS. With the ovarian cancer screening studies ongoing,[152] a similar opportunity will become available for early ovarian cancer. There is considerable scope for mathematical modelling to investigate hypotheses about biomarkers for the development of ovarian cancer.[153]

In Asian countries, there are a number of screening strategies for malignancies common in Asian populations. These include

- visual inspection for oral premalignancies,
- two stage screening (Epstein-Barr virus testing followed by endoscopy) for nasopharyngeal cancer,
- two stage screening (hepatitis testing together with family history of liver cancer followed by ultrasound) for liver cancer, and
- various approaches to screening and control of precursors of gastric cancer.

For many of these issues, simple analytic methods will sufficient. For some, however, explicit mathematical modelling of the disease process will be necessary. It is reasonable to expect

that problems of heterogeneity of progression will be common, and that some of the methods in this thesis will be relevant.

References

1. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 1979;**16**:439-54.

2. Hougaard P. Life Table Methods for Heterogeneous Populations - Distributions Describing the Heterogeneity. *Biometrika* 1984;**71**:75-83.

3. Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. *The Annals of Applied Probability* 1992;**2**:951-72.

4. Hougaard P. A Class of Multivariate Failure Time Distributions. *Biometrika* 1986;**73**:671-8.

5. Aalen OO. Heterogeneity in survival analysis. *Stat.Med.* 1988;**7**:1121-37.

6. Price DL,.Manatunga AK. Modelling survival data with a cured fraction using frailty models. *Stat.Med.* 2001;**20**:1515-27.

7. Vaida F,.Xu RH. Proportional hazards model with random effects. *Statistics in Medicine* 2000;**19**:3309-24.

8. Clayton DG. A Monte Carlo method for Bayesian inference in frailty models. *Biometrics* 1991;**47**:467-85.

9. Ripatti S, Gatz M, Pedersen NL, Palmgren J. Three-state frailty model for age at onset of dementia and death in Swedish twins. *Genet.Epidemiol.* 2003;**24**:139-49.

10. Hougaard P. Frailty models for survival data. *Lifetime Data Anal.* 1995;**1**:255-73.

11. Hougaard P, Myglegaard P, Borchjohnsen K. Heterogeneity Models of Disease Susceptibility, with Application to Diabetic Nephropathy. *Biometrics* 1994;**50**:1178-88.

12. Aalen OO,.Tretli S. Analyzing incidence of testis cancer by means of a frailty model. *Cancer Causes Control* 1999;**10**:285-92.

13. Hougaard P. Survival Models for Heterogeneous Populations Derived from Stable-Distributions. *Biometrika* 1986;**73**:386-96.

14. Longini IM,.Halloran ME. A frailty mixture model for estimating vaccine efficacy. *Applied Statistics-Journal of the Royal Statistical Society Series C* 1996;**45**:165-73.

15. Halloran ME, Longini IM, Struchiner CJ. Estimability and interpretation of vaccine efficacy using frailty mixing models. *American Journal of Epidemiology* 1996;**144**:83-97.

16. Boily MC, Masse BR, Desai K, Alary M, Anderson RM. Some important issues in the planning of phase III HIV vaccine efficacy trials. *Vaccine* 1999;**17**:989-1004.

17. Desai KN, Boily MC, Masse BR, Alary M, Anderson RM. Simulation studies of phase III clinical trials to test the efficacy of a candidate HIV-1 vaccine. *Epidemiol.Infect.* 1999;**123**:65-88.

18. Keiding N, Andersen PK, Klein JP. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Stat.Med.* 1997;**16**:215-24.

19. Begun AZ, Iachine IA, Yashin AI. Genetic nature of individual frailty: comparison of two approaches. *Twin.Res.* 2000;**3**:51-7.

20. Mack W, Langholz B, Thomas DC. Survival models for familial aggregation of cancer. *Environ.Health Perspect.* 1990;**87**:27-35.

21. Wienke A, Holm NV, Skytthe A, Yashin AI. The heritability of mortality due to heart diseases: a correlated frailty model applied to Danish twins. *Twin.Res.* 2001;**4**:266-74.

22. Yashin AI, Begun AZ, Iachine IA. Genetic factors in susceptibility to death: a comparative analysis of bivariate survival models. *J.Epidemiol.Biostat.* 1999;**4**:53-60.

23. Yashin AI, Iachine IA, Harris JR. Half of the variation in susceptibility to mortality is genetic: findings from Swedish twin survival data. *Behav.Genet.* 1999;**29**:11-9.

24. Yashin AI,.Iachine IA. Genetic analysis of durations: correlated frailty model applied to survival of Danish twins. *Genet.Epidemiol.* 1995;**12**:529-38.

25. Yashin AI,.Iachine IA. How frailty models can be used for evaluating longevity limits: taking advantage of an interdisciplinary approach. *Demography* 1997;**34**:31-48.

26. Zahl PH,.Harris JR. Cancer incidence for Swedish twins studied by means of bivariate frailty models. *Genet.Epidemiol.* 2000;**19**:354-65.

27. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J.Intern.Med.* 2002;**252**:247-54.

28. Siegmund K,.McKnight B. Modeling hazard functions in families. *Genet.Epidemiol.* 1998;**15**:147-71.

29. Siegmund KD, Todorov AA, Province MA. A frailty approach for modelling diseases with variable age of onset in families: the NHLBI Family Heart Study. *Stat.Med.* 1999;**18**:1517-28.

30. Klein JP. Semiparametric estimation of random effects using the Cox model based on the EM algorithm. *Biometrics* 1992;**48**:795-806.

31. Wassell JT,.Moeschberger ML. A bivariate survival model with modified gamma frailty for assessing the impact of interventions. *Stat.Med.* 1993;**12**:241-8.

32. Mahe C,.Chevret S. Estimating regression parameters and degree of dependence for multivariate failure time data. *Biometrics* 1999;**55**:1078-84.

33. Robertsson O,.Ranstam J. No bias of ignored bilaterality when analysing the revision risk of knee prostheses: Analysis of a population based sample of 44,590 patients with 55,298 knee prostheses from the national Swedish Knee Arthroplasty Register. *BMC.Musculoskelet.Disord.* 2003;**4**:1.

34. Liang KY, Self SG, Bandeen-Roche KJ, Zeger SL. Some recent developments for regression analysis of multivariate failure time data. *Lifetime Data Anal.* 1995;**1**:403-15.

35. Frydman H. Maximum-Likelihood Estimation in the Mover-Stayer Model. *Journal of the American Statistical Association* 1984;**79**:632-8.

36. Sampson M. A Markov-Chain Model for Unskilled Workers and the Highly Mobile. *Journal of the American Statistical Association* 1990;**85**:177-80.

37. Fuchs C,.Greenhouse JB. The EM algorithm for maximum likelihood estimation in the mover-stayer model. *Biometrics* 1988;**44**:605-13.

38. Swensen AR. On maximum likelihood estimation in the mover-stayer model. *Communications in Statistics-Theory and Methods* 1996;**25**:1717-28.

39. Blumen I, Kogan M, McCarthy P. Stochastic models for social processes. New York: John Wiley, 1955.

40. Addison JT,.Portugal P. Unemployment duration - Competing and defective risks. *Journal of Human Resources* 2003;**38**:156-91.

41. Davies RB, Elias P, Penn R. The Relationship Between A Husbands Unemployment and His Wifes Participation in the Labor-Force. *Oxford Bulletin of Economics and Statistics* 1992;**54**:145-71.

42. Kettunen J. Labour mobility of unemployed workers. *Regional Science and Urban Economics* 2002;**32**:359-80.

43. Lavoie M,.Finnie R. The occupational dynamics of recent Canadian engineering graduates inside and outside the bounds of technology. *Research Policy* 1998;**27**:143-58.

44. Light A,.McGarry K. Job change patterns and the wages of young men. *Review of Economics and Statistics* 1998;**80**:276-86.

45. Baydar N,.White M. A method for analyzing backward recurrence time data on residential mobility. *Sociol.Methodol.* 1988;**18**:105-35.

46. Concord CMS. A Mover Stayer Approach to Residential-Mobility of the Aged. *Tijdschrift Voor Economische en Sociale Geografie* 1984;**75**:249-62.

47. Chaterjee R,.Ramaswamy V. An extended mover-stayer model for diagnosing the dynamics of trial and repeat for a new brand. *Applied Stochastic Models and Data Analysis* 1996;**12**:165-78.

48. Dutta J, Sefton JA, Weale MR. Income distribution and income dynamics in the United Kingdom. *Journal of Applied Econometrics* 2001;**16**:599-617.

49. Yamaguchi K. Mover-stayer models for analyzing event nonoccurrence and event timing with time-dependent covariates: An application to an analysis of remarriage. *Sociological Methodology, Vol.28 1998* 1998;**28**:327-61.

50. Fule E. Changes on the Czech political scene. *Electoral Studies* 1997;**16**:341-7.

51. Morgan TM, Aneshensel CS, Clark VA. Parameter-Estimation for Mover-Stayer Models - Analyzing Depression Over Time. *Sociological Methods & Research* 1983;**11**:345-66.

52. Pasqualucci C, Rava L, Rossi C, Schinaia G. Estimating the size of the HIV/AIDS epidemic: Complementary use of the empirical Bayesian Back-Calculation and the Mover-Stayer model for gathering the largest amount of information. *Simulation* 1998;**71**:213-27.

53. Rossi C. A Stochastic Mover Stayer Model for An Hiv Epidemic. *Mathematical Biosciences* 1991;**107**:521-45.

54. Rossi C,.Schinaia G. The mover-stayer model for the HIV/AIDS epidemic in action. *Interfaces* 1998;**28**:127-43.

55. Schinaia G. Modeling the HIV/AIDS epidemic via survivor functions. *European Journal of Epidemiology* 2000;**16**:573-9.

56. Chen HH, Duffy SW, Tabar L. A mover-stayer mixture of Markov chain models for the assessment of dedifferentiation and tumour progression in breast cancer. *Journal of Applied Statistics* 1997;**24**:265-78.

57. Albert PS. A mover-stayer model for longitudinal marker data. *Biometrics* 1999;**55**:1252-7.

58. Morrison AS. Screening in chronic disease. New York: Oxford University Press, 1985.

59. Flanders WD,.Longini IM, Jr. Estimating benefits of screening from observational cohort studies. *Stat.Med.* 1990;**9**:969-80.

60. Feuer EJ, Mariotto A, Merrill R. Modeling the impact of the decline in distant stage disease on prostate carcinoma mortality rates. *Cancer* 2002;**95**:870-80.

61. van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JD. Cost-effectiveness of cervical cancer screening: comparison of screening policies. *J.Natl.Cancer Inst.* 2002;**94**:193-204.

62. Tabar L, Fagerberg G, Chen HH, Duffy SW, Gad A. Tumour development, histology and grade of breast cancers: prognosis and progression. *Int.J.Cancer* 1996;**66**:413-9.

63. Brinton LA,.Fraumeni JF, Jr. Epidemiology of uterine cervical cancer. *J.Chronic.Dis.* 1986;**39**:1051-65.

64. van Oortmarssen GJ,.Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. *Br.J.Cancer* 1991;**64**:559-65.

65. Kronborg O. Colon polyps and cancer. *Endoscopy* 2004;**36**:3-7.

66. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS *et al.* Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N.Engl.J.Med.* 1993;**329**:1977-81.

67. Laara E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet* 1987;**1**:1247-9.

68. Ernster VL, Barclay J, Kerlikowske K, Grady D, Henderson IC. Incidence of and treatment for ductal carcinoma in situ of the breast. *Jama-Journal of the American Medical Association* 1996;**275**:913-8.

69. Ernster VL,.Barclay J. Increases in ductal carcinoma in situ (DCIS) of the breast in relation to mammography: a dilemma. *J.Natl.Cancer Inst.Monogr* 1997;151-6.

70. Feig SA. Ductal carcinoma in situ - Implications for screening mammography. *Radiologic Clinics of North America* 2000;**38**:653-+.

71. Ernster VL, Ballard-Barbash R, Barlow WE, Zheng YY, Weaver DL, Cutter G *et al.* Detection of ductal carcinoma in situ in women undergoing screening mammography. *Journal of the National Cancer Institute* 2002;**94**:1546-54.

72. Kerlikowske K, Barclay J, Grady D, Sickles EA, Ernster V. Comparison of risk factors for ductal carcinoma in situ and invasive breast cancer. *Journal of the National Cancer Institute* 1997;**89**:76-82.

73. Weiss HA, Brinton LA, Brogan D, Coates RJ, Gammon MD, Malone DE *et al.* Epidemiology of in situ and invasive breast cancer in women aged under 45. *British Journal of Cancer* 1996;**73**:1298-305.

74. Welch HG,.Black WC. Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: How much more breast cancer can we find? *Annals of Internal Medicine* 1997;**127**:1023-8.

75. Nielsen M, Jensen J, Andersen J. Precancerous and Cancerous Breast-Lesions During Lifetime and at Autopsy - A Study of 83 Women. *Cancer* 1984;**54**:612-5.

76. Nielsen M, Thomsen JL, Primdahl S, Dyreborg U, Andersen JA. Breast-Cancer and Atypia Among Young and Middle-Aged Women - A Study of 110 Medico-Legal Autopsies. *British Journal of Cancer* 1987;**56**:814-9.

77. Betsill WL, Jr., Rosen PP, Lieberman PH, Robbins GF. Intraductal carcinoma. Long-term follow-up after treatment by biopsy alone. *JAMA* 1978;**239**:1863-7.

78. Rosen PP, Braun DW, Jr., Kinne DE. The clinical significance of pre-invasive breast carcinoma. *Cancer* 1980;**46**:919-25.

79. Page DL, Dupont WD, Rogers LW, Landenberger M. Intraductal Carcinoma of the Breast - Follow-Up After Biopsy Only. *Cancer* 1982;**49**:751-8.

80. Page DL, Dupont WD, Rogers LW, Jensen RA, Schuyler PA. Continued Local Recurrence of Carcinoma 15-25 Years After A Diagnosis of Low-Grade Ductal Carcinoma In-Situ of the Breast Treated Only by Biopsy. *Cancer* 1995;**76**:1197-200.

81. Eusebi V, Foschini MP, Cook MG, Berrino F, Azzopardi JG. Long-Term Follow-Up of Insitu Carcinoma of the Breast with Special Emphasis on Clinging Carcinoma. *Seminars in Diagnostic Pathology* 1989;**6**:165-73.

82. Evans AJ, Pinder SE, Ellis IO, Wilson ARM. Screen detected ductal carcinoma in situ (DCIS): overdiagnosis or an obligate precursor of invasive disease? *Journal of Medical Screening* 2001;**8**:149-51.

83. Duffy SW, Tabar L, Vitak B, Day NE, Smith RA, Chen HHT *et al.* The relative contributions of screen-detected in situ and invasive carcinomas in reducing mortality from the disease. *European Journal of Cancer* 2003;**(In press)**.

84. Tabar L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF *et al.* The Swedish two-county trial twenty years later - Updated mortality results and new insights from long-term follow-up. *Radiologic Clinics of North America* 2000;**38**:625-+.

85. Chamberlain J, Moss SM, Kirkpatrick AE, Michell M, Johns L. National-Health-Service Breast Screening-Program Results for 1991-2. *British Medical Journal* 1993;**307**:353-6.

86. Fracheboud J, de Koning HJ, Beemsterboer PMM, Boer R, Hendriks JHCL, Verbeek ALM *et al.* Nation-wide breast cancer screening in the Netherlands: Results of initial and subsequent screening 1990-1995. *International Journal of Cancer* 1998;**75**:694-8.

87. Robinson JI, Crane CEB, King JM, Scarce DI, Hoffmann CEJ. The South Australian breast X-ray service: Results from a statewide mammographic screening programme. *British Journal of Cancer* 1996;**73**:837-42.

88. Dershaw DD, Loring L, True SJ, Momrow DC, Liberman L, Morris EA. A comparison of screening mammography results from programs for women of different socioeconomic status. *Cancer* 1998;**82**:1692-7.

89. Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer part I: tumour attributes and the preclinical screen-detectable phase.

Journal of Epidemiology and Biostatistics 1997;**2**:9-23.

90. Cox DR, Miller HD. The theory of stochastic process. London: Methuen, 1965.

91. Vezeridis MP,.Bland KI. Management of Ductal Carcinoma In-Situ. *Surgical Oncology-Oxford* 1994;**3**:309-25.

92. McCready DR. Ductal carcinoma in situ. In Souhami RL, Hohenberger p, Horiot JC, eds. *Oxford Textbook of Oncology*, pp 1717-24. Oxford: Oxford University Press, 2002.

93. Silverstein MJ, Lagios MD, Craig PH, Waisman JR, Lewinsky BS, Colburn WJ *et al.* A prognostic index for ductal carcinoma in situ of the breast. *Cancer* 1996;**77**:2267-74.

94. Frykberg ER,.Bland KI. Overview of the Biology and Management of Ductal Carcinoma In- Situ of the Breast. *Cancer* 1994;**74**:350-61.

95. Bertario L, Russo A, Sala P, Pizzetti P, Ballardini G, Andreola S *et al.* Predictors of metachronous colorectal neoplasms in sporadic adenoma patients. *Int.J.Cancer* 2003;**105**:82-7.

96. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M *et al.* Genetic alterations during colorectal-tumor development. *N.Engl.J.Med.* 1988;**319**:525-32.

97. Winawer SJ, O'Brien MJ, Waye JD, Kronborg O, Bond J, Fruhmorgen P *et al.* Risk and surveillance of individuals with colorectal polyps. Who Collaborating Centre for the Prevention of Colorectal Cancer. *Bull.World Health Organ* 1990;**68**:789-95.

98. Yood MU, Oliveria S, Boyer JG, Wells K, Stang P, Johnson CC. Colon polyp recurrence in a managed care population. *Arch.Intern.Med.* 2003;**163**:422-6.

99. Cole P,.Morrison AS. Basic issues in population screening for cancer. *J.Natl.Cancer Inst.* 1980;**64**:1263-72.

100. Kuffer R,.Lombardi T. Premalignant lesions of the oral mucosa. A discussion about the place of oral intraepithelial neoplasia (OIN). *Oral Oncol.* 2002;**38**:125-30.

101. Laara E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet* 1987;**1**:1247-9.

102. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history

of untreated colonic polyps. *Gastroenterology* 1987;**93**:1009-13.

103. Kudo S, Kashida H, Tamura T. Early colorectal cancer: flat or depressed type. *J.Gastroenterol.Hepatol.* 2000;**15 Suppl**:D66-D70.

104. Chen CD, Yen MF, Wang WM, Wong JM, Chen TH. A case-cohort study for the disease natural history of adenoma-carcinoma and de novo carcinoma and surveillance of colon and rectum after polypectomy: implication for efficacy of colonoscopy. *Br.J.Cancer* 2003;**88**:1866-73.

105. Aalen OO. Effects of frailty in survival analysis. *Stat.Methods Med.Res.* 1994;**3**:227-43.

106. Chen TH, Kuo HS, Yen MF, Lai MS, Tabar L, Duffy SW. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics* 2000;**56**:167-72.

107. Chen TH, Yen MF, Shiu MN, Tung TH, Wu HM. Stochastic model for non-standard case-cohort design. *Stat.Med.* 2004;**23**:633-47.

108. Duffy SW, Chen HH, Tabar L, Day NE. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Stat.Med.* 1995;**14**:1531-43.

109. Myers ER, McCrory DC, Subramanian S, McCall N, Nanda K, Datta S *et al.* Setting the target for a better cervical screening test: characteristics of a cost-effective test for cervical neoplasia screening. *Obstet.Gynecol.* 2000;**96**:645-52.

110. Pokorski RJ. Long-term morbidity and mortality risk in Japanese insurance applicants with chronic hepatitis C virus infection. *J.Insur.Med.* 2001;**33**:12-36.

111. SAS Institute inc. SAS/IML (R) Software: Changes and enhancements through Release 6.11. SAS Institute Inc., 1998.

112. Tabar L, Chen HH, Duffy SW, Yen MF, Chiang CF, Dean PB *et al.* A novel method for prediction of long-term outcome of women with T1a, T1b, and 10-14 mm invasive breast cancers: a prospective study. *Lancet* 2000;**355**:429-33.

113. Peacock C, Given-Wilson RM, Duffy SW. Mammographic casting-type calcification associated with small screen-detected invasive breast cancers: is this a reliable prognostic indicator? *Clin.Radiol.* 2004;**59**:165-70.

114. Thurfjell E, Thurfjell MG, Lindgren A. Mammographic finding as predictor of survival in 1-9 mm invasive breast cancers. worse prognosis for cases presenting as calcifications alone. *Breast Cancer Res.Treat.* 2001;**67**:177-80.

115. Zunzunegui RG, Chung MA, Oruwari J, Golding D, Marchant DJ, Cady B. Casting-type calcifications with invasion and high-grade ductal carcinoma in situ: a more aggressive disease? *Arch.Surg.* 2003;**138**:537-40.

116. Seymour HR, Cooke J, Given-Wilson RM. The significance of spontaneous resolution of breast calcification. *Br.J.Radiol.* 1999;**72**:3-8.

117. Tabar L, Yen MF, Vitak B, Tony Chen HH, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet* 2003;**361**:1405-10.

118. Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol.Clin.North Am.* 1992;**30**:187-210.

119. Tabar L, Chen THH, Yen MF, Tot T, Tung TH, Chen LS *et al.* Mammographic tumor features can reliably predict the long-term outcome of women with 1-14 mm invasive breast cancer: suggestions for revision of current therapeutic practice and the TNM classification system. *Cancer* 2004;**In press**.

120. McKee GT, Tildsley G, Hammond S. Cytologic diagnosis and grading of ductal carcinoma in situ. *Cancer* 1999;**87**:203-9.

121. Cervical cancer screening programs. I. Epidemiology and natural history of carcinoma of the cervix. [Review] [41 refs]. *Canadian Medicine Association Journal* 1976;1003-12.

122. Bourne RG,.Grove WD. Invasive carcinoma of the cervix in Queensland. Change in incidence and mortality, 1959-1980. *Med.J.Aust.* 1983;**1**:156-8.

123. Hakama M, Chamberlain J, Day NE, Miller AB, Prorok PC. Evaluation of screening programmes for gynaecological cancer. *Br.J.Cancer* 1985;**52**:669-73.

124. Duguid HL, Duncan ID, Currie J. Screening for cervical intraepithelial neoplasia in Dundee and Angus 1962-81 and its relation with invasive cervical cancer. *Lancet* 1985;**2**:1053-6.

125. Pettersson F, Bjorkholm E, Naslund I. Evaluation of screening for cervical cancer in

Sweden: trends in incidence and mortality 1958-1980. *Int.J.Epidemiol.* 1985;**14**:521-7.

126. Laara E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet* 1987;**1**:1247-9.

127. Hakama M,.Rasanen-Virtanen U. Effect of a mass screening program on the risk of cervical cancer. *Am.J.Epidemiol.* 1976;**103**:512-7.

128. Day NE. Effect of cervical cancer screening in Scandinavia. *Obstet.Gynecol.* 1984;**63**:714-8.

129. Johannesson G, Geirsson G, Day N. The effect of mass screening in Iceland, 1965-74, on the incidence and mortality of cervical carcinoma. *Int.J.Cancer* 1978;**21**:418-25.

130. Ebeling K,.Nischan P. Assessing the effectiveness of a cervical cancer screening program in the German Democratic Republic. *Int.J.Technol.Assess.Health Care* 1987;**3**:137-47.

131. Duguid HL, Duncan ID, Currie J. Screening for cervical intraepithelial neoplasia in Dundee and Angus 1962-81 and its relation with invasive cervical cancer. *Lancet* 1985;**2**:1053-6.

132. Gardner JW,.Lyon JL. Efficacy of cervical cytologic screening in the control of cervical cancer. *Prev.Med.* 1977;**6**:487-99.

133. Miller AB, Lindsay J, Hill GB. Mortality from cancer of the uterus in Canada and its relationship to screening for cancer of the cervix. *Int.J.Cancer* 1976;**17**:602-12.

134. Peto J, Gilham C, Fletcher O, Matthews FE. The cervical cancer epidemic that screening has prevented in the UK. *Lancet* 2004;**364**:249-56.

135. Vizcaino AP, Moreno V, Bosch FX, Munoz N, Barros-Dios XM, Borras J *et al.* International trends in incidence of cervical cancer: II. Squamous-cell carcinoma. *Int.J.Cancer* 2000;**86**:429-35.

136. Parkin DM, Nguyen-Dinh X, Day NE. The impact of screening on the incidence of cervical cancer in England and Wales. *Br.J.Obstet.Gynaecol.* 1985;**92**:150-7.

137. Lee SS, Collins RJ, Pun TC, Cheng DK, Ngan HY. Conservative treatment of low grade squamous intraepithelial lesions (LSIL) of the cervix. *Int.J.Gynaecol.Obstet.* 1998;**60**:35-40.

138. Raffle AE, Alden B, Mackenzie EF. Detection rates for abnormal cervical smears: what are we screening for? *Lancet* 1995;**345**:1469-73.

139. Raffle AE, Alden B, Quinn M, Babb PJ, Brett MT. Outcomes of screening to prevent cancer: analysis of cumulative incidence of cervical abnormality and modelling of cases and deaths prevented. *BMJ* 2003;**326**:901.

140. Shapiro S, Goldberg JD, Hutchison GB. Lead time in breast cancer detection and implications for periodicity of screening. *Am.J.Epidemiol.* 1974;**100**:357-66.

141. Prorok PC. Mathematical models and natural history in cervical cancer screening. *IARC Sci.Publ.* 1986;185-98.

142. Knox EG. A simulation system for screening procedures: computer simulations of cervical cytology screening programmed. In McLachlan EG, ed. *Problems and progress in Medical Care*, pp 30-55. London: Oxford University Press, 1973.

143. Knox EG. Ages and frequencies for cervical cancer screening. *Br.J.Cancer* 1976;**34**:444-52.

144. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. Model building on the basis of Dutch cervical cancer screening data. *Maturitas* 1985;**7**:11-20.

145. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. *Comput.Methods Programs Biomed.* 1985;**20**:79-93.

146. Parkin DM. A computer simulation model for the practical planning of cervical cancer screening programmes. *Br.J.Cancer* 1985;**51**:551-68.

147. Gustafsson L,.Adami HO. Natural history of cervical neoplasia: consistent results obtained by an identification technique. *Br.J.Cancer* 1989;**60**:132-41.

148. van Oortmarssen GJ,.Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. *Br.J.Cancer* 1991;**64**:559-65.

149. Bos AB, van Ballegooijen M, van Oortmarssen GJ, van Marle ME, Habbema JD, Lynge E. Non-progression of cervical intraepithelial neoplasia estimated from population-screening data. *Br.J.Cancer* 1997;**75**:124-30.

150. Sasieni P, Adams J, Cuzick J. Benefit of cervical screening at different ages: evidence from the UK audit of screening histories. *Br.J.Cancer* 2003;**89**:88-93.

151. Canfell K, Barnabas R, Patnick J, Beral V. The predicted effect of changes in cervical screening practice in the UK: results from a modelling study. *Br.J.Cancer* 2004;**91**:530-6.

152. Jacobs IJ,.Menon U. Progress and challenges in screening for early detection of ovarian cancer. *Mol.Cell Proteomics.* 2004;**3**:355-66.

153. Skates SJ, Horick N, Yu Y, Xu FJ, Berchuck A, Havrilesky LJ *et al.* Preoperative Sensitivity and Specificity for Early-Stage Ovarian Cancer When Combining Cancer Antigen CA-125II, CA 15-3, CA 72-4, and Macrophage Colony-Stimulating Factor Using Mixtures of Multivariate Normal Distributions. *J.Clin.Oncol.* 2004.

Appendix 1 SAS macro for three-state model conjugated with frailty term applying to one transition rate between two states

```
%macro G_pf1;

  %macro G_prob11;
    p11= Exp(-(b1*predt));
  %mend G_prob11;

  %macro G_prob12;
    start fun12(s) global(predt,b1,b20,theta,delta);
      v12=b1*exp(-b1*s)*( theta / (theta+b20*(predt-s)) )**delta ;
      return(v12);
    finish fun12;
    call quad(p12,"fun12",indext);
  %mend G_prob12;

  %macro G_prob13;
    start fun13(s) global(predt,b1,b20,theta,delta,xs);
      xs=s;
      indext23=J(2,1,0); indext23[2]=predt-xs;

      start fun23(u) global(predt,b1,b20,theta,delta,xs);

        v23=( theta / (theta+b20*u )) **delta  *  (delta*b20/(theta+b20*u)) ;
        return(v23);
      finish fun23;
      call quad(p23,"fun23",indext23);
      v13=b1*exp(-b1*s)*p23;
      return(v13);
    finish fun13;
    call quad(p13,"fun13",indext);
  %mend G_prob13;

%mend G_pf1;


%macro G_pf2;

  predt=tt[i,1];
  indext=J(2,1,0);  indext[2]=predt;
  b1=h[1]; b20=h[2]; theta=h[3]; delta=h[4];
  %G_prob11; %G_prob12; %G_prob13;
  p0=p11;
  p1=p12;
  p2=p13;

%mend G_pf2;


%macro G_para;
  h0={  0.002  0.3   1   1   };
  con={ 0     0     0   0   ,
        .     .     .   .   };
```

```
%mend G_para;

%macro CP_pf1;
  %macro CP_pr11;
    p11= Exp(-(b1*predt));
  %mend CP_pr11;

  %macro CP_pr12;
    start fun12(s) global(predt,b1,b20,tho,fi,ita);
      v12=b1*exp(-b1*s + tho* ( (1+b20*(predt-s)/fi)**(-ita) ) - tho  ) ;
      return(v12);
    finish fun12;
    call quad(p12,"fun12",indext);

  %mend CP_pr12;

  %macro CP_pr13;
    start fun13(s) global(predt,b1,b20,tho,fi,ita,xs);
      xs=s;
      indext23=J(2,1,0); indext23[2]=predt-xs;
      start fun23(u) global(predt,b1,b20,b4,tho,fi,ita,xs);
        v23=( tho*ita*b20 ) / (fi * ( (1+b20*u/fi) ** (ita+1))  ) *
            exp( tho* ( (1+b20*u/fi)**(-ita) ) - tho );
        return(v23);
      finish fun23;
      call quad(p23,"fun23",indext23);
      v13=b1*exp(-b1*s)*p23;
      return(v13);
    finish fun13;
    call quad(p13,"fun13",indext);
  %mend CP_pr13;

%mend CP_pf1;


%macro CP_pf2;
  predt=tt[i,1];
  indext=J(2,1,0);  indext[2]=predt;
  b1=h[1]; b20=h[2]; tho=h[3]; fi=h[4];  ita=h[5];
  %cp_pr11; %cp_pr12; %cp_pr13;
  p0=p11;
  p1=p12;
  p2=p13;

%mend CP_pf2;


%macro CP_para;
  h0={  0.002  0.3   0.4  1   0.9 };
  con={ 0     0    0   0  0   ,
        .     .    .   .  .  };
%mend CP_para;



%macro MS_pf1;

  start pmtrxa(h,t);
```

```
    Q=J(3,3,0);
    Q[1,1]=-h[1];
    Q[1,2]= h[1];
    Q[2,2]=-h[2];
    Q[2,3]= h[2];
    A=teigvec(Q);
    v=teigval(Q);
    D=diag(exp(v[,1]#t));
    P=A*D*inv(A);
    return(P);
  finish pmtrxa;

  start pmtrxb(h,t);
    Q=J(2,2,0);
    Q[1,1]=-h[1];
    Q[1,2]= h[1];
    A=teigvec(Q);
    v=teigval(Q);
    D=diag(exp(v[,1]#t));
    P=A*D*inv(A);
    return(P);
  finish pmtrxb;

%mend MS_pf1;


%macro MS_pf2;

  Pa=pmtrxa(h,tt[i,1]);
  Pb=pmtrxb(h,tt[i,1]);
  mover=h[3];  stayer=1-mover;

  p0=mover*pa[1,1]+stayer*pb[1,1];
  p1=mover*pa[1,2]+stayer*pb[1,2];
  p2=mover*pa[1,3];

%mend MS_pf2;

%macro MS_para;

   h0={ 0.005   0.015  0.6};
  con={ 1.e-5   1.e-5  0  ,
       1      1     1  };

%mend MS_para;


%macro hetero(type, tt, para);

title '3 state, Excluding hyperplasia';
title2 "Heterogeneity model, Type: &tt";

 proc iml;
  use mydata;
  read all into tt[colname=col];
  m=nrow(tt);

 %&type._pf1;
```

```
 start f_logL(h) global(tt,m &para);
    sum=0;
    do i=1 to m;

    propor1=305/10496;
    propor2=300/2627;
    propor3=116/760;

    %&type._pf2;

    px0=p0*propor1/(propor1*p0+propor2*p1+propor3*p2);
    px1=p1*propor2/(propor1*p0+propor2*p1+propor3*p2);
    px2=p2*propor3/(propor1*p0+propor2*p1+propor3*p2);

         if tt[i,2]=0 then sum=sum+tt[i,3]*log(px0);
     else if tt[i,2]=1 then sum=sum+tt[i,3]*log(px1);
     else if tt[i,2]=2 then sum=sum+tt[i,3]*log(px2);
     end;
    return(sum);
 finish f_logL;

 %&type._para;

optn={1 2};
 call nlpnra(rc,xres,"f_logL",h0,optn,con);
  estimate=xres`;
 call nlpfdd(f,g,hes2,"f_logL",estimate);
  cov=-inv(hes2);
  print "Variance-Covariance Matrix";
  print cov;
  prob=.05;
  norqua=probit(1-prob/2);
  stderr=sqrt(vecdiag(cov));
  lowbound=estimate-norqua*stderr;
   upbound=estimate+norqua*stderr;
  print "Asymptotic 95% Confidence Interval";
  print lowbound estimate upbound stderr;
  print rc;
quit;

%mend hetero;


%hetero(type= G ,
       tt=Gamma-distributed,
       para= %str(,) predt %str(,) b1 %str(,) b20 %str(,) theta %str(,) delta);

%hetero(type= CP ,
       tt=compound Poisson_distributed,
       para= %str(,) predt %str(,) b1 %str(,) b20 %str(,) tho %str(,) fi %str(,) ita);

%hetero(type= MS ,
       tt=mover-stayer model,
       para=);
```