ESTIMATING PARAMETERS OF A MICROSIMULATION MODEL FOR BREAST CANCER SCREENING USING THE SCORE FUNCTION METHOD

Sita Y.G.L. Tan (1)
Gerrit J. van Oortmarssen (1)
Nanda Piersma (2)

(1) Department of Public Health, Faculty of Medicine and Health Sciences, Erasmus University, Rotterdam, The Netherlands.
(2) Econometric Institute, Erasmus University, Rotterdam, The Netherlands.

Econometric Institute Report EI2000-35/A

Abstract

In developing decision-making models for the evaluation of medical procedures, the model parameters can be estimated by fitting the model to data observed in trial studies. For complex models that are implemented by discrete event simulation (microsimulation) of individual life histories, the Score Function (SF) method can potentially be an appropriate approach for such estimation exercises. We test this approach for a microsimulation model of screening for cancer that is fitted to data from the HIP randomized trial for early detection of breast cancer. Comparison of the parameter values estimated by the SF method and the analytical solution shows that method performs well on this simple model. The precision of the estimated parameter values depends (as expected) on the size of the simulation (number of life histories), and on the number of parameters estimated. Using analytical representations for parts of the microsimulation model can increase the precision in the estimation of the remaining parameters. Compared to the Nelder and Mead Simplex method which is often used in stochastic simulation because of its ease of implementation, the SF method is clearly more efficient (ratio computer time: precision of estimates). The additional analytical investment needed to implement the method in an (existing) simulation model may well be worth the effort.

Keywords: Medical decision making, (Micro) simulation, Score function, Quasi-Newton method.

AMS(MOS) Classification: 92C60, 90C53, 68U20

1. Introduction

Microsimulation models are increasingly used in epidemiology and public health for analyzing data from disease control projects and for evaluating the costs and effects of control policies. Microsimulation of disease control is a specific type of discrete event simulation in which individual life histories of fictitious persons are generated, including disease(s) in the persons and the effects of interventions, such as early detection (screening) and prevention and control of transmission of infectious diseases. These models have many parameters, and only part of these parameters can be quantified directly on the basis of existing knowledge. Inferences for other parameters may be obtained by optimizing the goodness of fit between the model-simulated data and data from intervention trials. Examples of these models are the MISCAN model for breast cancer screening [1] and for colorectal cancer screening [2]. These microsimulation models, which can be regarded as extended and more realistic alternatives for simpler numerical-analytical models [3], have been used in decision making on implementation of a national mass screening program for breast cancer in The Netherlands and in evaluation of the results of screening programs.

An essential step in the quantification of the models is the estimating parameter values from available screening trial data. The stochastic output and the long duration of microsimulation runs make that finding precise parameter estimates may be very time-consuming. The same applies to finding optimal disease control policies such as policies that give the highest gain in life years for a certain budget. In this paper we will concentrate on finding the parameter estimates of microsimulation models. A search method for the best parameter estimates requires running the microsimulation for different parameter values and comparing the performance of the simulation to that of recorded trial data. Typically fitting the parameters of a microsimulation model results in an optimization problem where the objective function (the goodness of fit of the model) has the following characteristics:

1. Calculation of the objective function is very expensive or time-consuming.
2. Exact first partial derivatives of this function cannot be calculated.
3. Numerical approximation of the gradient of the objective function is impracticably expensive or slow.

An example of a direct search method, i.e. a method that only uses the value of the objective function [4] is the Nelder and Mead Simplex (NMSM) algorithm [5], because it considers the (micro)simulation model as a black box. Therefore, this method can be applied to any (micro)simulation model or optimization problem. In the NMSM approach, each step in the optimization algorithms is based on output from a number of simulation runs with a large number of life histories. The basic NMSM is based on ranking of vertices of the simplex with respect to model results, and it will often perform quite well in locating the (broad) neighborhood of the optimum. However, it is seriously hindered by noise in finding

precise optimal parameter values. A modified “adaptive version” of the simplex algorithm, in which the number of life histories simulated is increased when noise dominates the differences between simulation output between vertices, is more efficient than the basic algorithm [6].

Other approaches that consider the model to be a black box that does not give information on gradients or higher order derivatives are the Response Surface methodology [7],[8] and the Stochastic Approximation method [9]. These methods use evaluations of the objective function of a simulation model for different parameter values to calculate estimates of derivatives that are then used in optimization routines.

Opening the black box to obtain information during the (micro-)simulation or even adapting the simulation procedure may give more precise estimates of the objective function and its derivatives. Several approaches have been proposed, such as the quasi-gradient method that is usually combined with a stochastic approximation type of optimization method. Here the specifics of the microsimulation are important for the estimation of the derivatives.

Another approach to (quasi) gradient optimization uses the Score Function (SF) method [10],[11]. In this optimization approach first a single (micro-)simulation run is performed, and then the score function is used to obtain successive estimates of the gradient and Hessian for different parameter values. In this way, the sample trajectories generated during the simulation run are used to model the objective function and its derivatives as a continuous deterministic function of the model parameters. Therefore, the model parameters can be optimized with a standard (quasi) Newton optimization method.

In the context of estimating parameters of disease control microsimulation models from empirical data, the Score Function method calculates a goodness of fit *and* its derivatives on the basis of only one (large) sample of simulated life histories that are generated for a certain "reference" value of the model parameters. The SF method can under mild conditions be added to an existing microsimulation model and will in general not change the generation of life histories.

We will investigate the use of the SF method for estimating parameters of a microsimulation version of the simplified analytical model for cancer screening proposed by Day and Walter [3],[12], using the same data set as they did from the HIP randomized controlled trial of breast cancer screening. The SF method requires that individual life histories are independent, which is true for this cancer-screening model. The efficiency of optimization using the SF-method is compared to that of the adaptive version of simplex (NMSM) method [6] by gauging the precision of the estimated optimum and the computational effort required.

Our tests for the simple breast cancer model will demonstrate that the SF method does indeed show a much better balance between precision of the parameter estimates and computational effort if compared to either the basic or the adaptive version of the NMSM. An additional advantage of the SF method is

that the outcome regarding gradients and Hessian matrices for the best-fitting parameter value can be used to estimate an (approximate) confidence region for the parameters. We will discuss the potential use of the SF method for more comprehensive model for breast cancer screening and for other medical decision making models.

2. Material and methods

The score function (SF) method is tested for a microsimulation model that is an implementation of a simple model for the detection of early stages of breast cancer by mass screening proposed by Walter and Day [3].In this model the preclinical disease consists of only one stage, the detectable pre-clinical period (DPCP), in which the cancer is already detectable by a screening test but does not cause serious symptoms. The DPCP starts when the tumor becomes detectable by the screening test used and ends when the tumor is detected clinically on the basis of symptoms. The model predicts the number of cancers detected by screening and the numbers of interval cancers, i.e. cancers that are missed by screening and detected clinically on the basis of symptoms. The model can be used to predict the number of cancers detected by screening for different screening strategies, or to estimate values of model parameters that can not be observed directly from numbers of screen detected and clinically diagnosed cancers in a trial study.

2.1 Material

The data used in this paper is from the first randomized controlled trial of breast cancer, the Health Insurance Plan of Greater New York (HIP) study. The aim of the HIP study, which started in 1963, was to establish whether early detection and treatment would result in a reduction of breast cancer mortality. The study involved approximately 62000 women between age 40 and 64 who were randomly divided over a study group and a control group of equal size. The women in the study group were invited to a first screening, and the 20166 (65%) participants of the first screening were offered three more screenings with intervals of one year [13]. The intervals after the screenings are called follow up intervals where a woman can only be diagnosed on the basis of symptoms. If a woman does not show for a screening the follow up intervals after the following screenings all relate to the last attended screening.

We consider the data for the first five years of the HIP study, because these were also used by Walter & Day for estimating a numerical version of the model [3],[12], see Figure 1. In this paper, the total number of screenings is denoted by S and the total number of follow-up intervals by I. In each of the

18 categories shown in the figure the number of women present in the category and the number of cancers observed have been registered in the HIP study. For a screening category G_s (s=1,..,S) the number of women who attended the screening and the number of cancers detected are recorded. For an interval group $G_{r,i}$ (r=1,..,S and i=r,..,I) where r indicates the number of screenings attended, the number of women who are present in this group (who were not detected in the previous screening or did not show for this screening) and the number of diagnosed cancers are recorded. At the initial screening test, breast cancer was detected in 55 of the 20166 women examined. In the year between the first and second screening, breast cancer was diagnosed in 13 women who had a negative test result at the first screening. These interval cancers were either fast growing tumors that were not yet detectable at the screening, or had been missed because of the limited sensitivity of the screening procedure. In the second screening, 15936 women participated and 32 cancers were detected. The detection rates at subsequent screenings are lower than at the first screening because the cancers with a long detectable phase can be detected at multiple screening instances. If the sensitivity of the test is good enough the cancers will be detected at the first screening occasion causing a lower detection rate at later screenings. The incidence rates for women who did not participate in a screening round is expected to increase with time since the last screening, and to return to the level for the situation without screening. These data are therefore considered separately in the analysis. Data for women who participated irregularly in the screenings have been neglected in the analysis.

==Insert Figure 1==

2.2 Microsimulation model

We consider the following microsimulion model, which is based on the simple one-stage breast cancer model as proposed Walter and Day [3]. The incidence of breast cancer is modeled by the hazard rate J $\in$ (0,1). If a cancer occurs then the starting time and the length of the DPCP completely determine the disease process.

Define the stochastic variable Y_l as the sojourn time (duration) of the DPCP and assume that the sojourn time has an exponential distribution function:

$$f_l(y) = l \cdot \exp(-l \cdot y) \qquad y \geq 0 \qquad (1)$$

where l is the sojourn time parameter.

Let t_1 and t_S be the points in time at which the first and the last screening take place respectively and t_{FU} as the point in time at which the follow-up period after the last screening ends (Figure 1). We only simulate relevant histories, i.e. life histories that can either be detected by screening or diagnosed as interval cancer before the end of the follow-up period. For given length y of the DPCP, these life histories all have a DPCP that starts within the interval $[t_1 - y, M(y) - y]$, where

$$M(y) = \max\{t_S - t_1 - y, t_{FU} - t_1\} \tag{2}$$

Define the stochastic variable X as the starting point of a relevant DPCP that, for given y, is uniformly distributed on $[t_1 - y, M(y) - y]$. The conditional probability density function (pdf) $k(x \mid y)$ is given by

$$k(x \mid y) = 1/(M(y) - t_1) \quad x \in [t_1 - y, M(y) - y] \tag{3}$$

The probability that a screening test will detect a cancer, depends on the incidence (hazard) rate *J*, the false-negative rate of the screening test (i.e. the probability that the screening test does not identify a cancer in the DPCP, denoted by b) and the sojourn time of the DPCP (parameter l). The model thus has three parameters that will be denoted by the parameter vector $v = (l, b, J)$.

Define the stochastic variable $A_{r,b}$ identifying the category in Figure 1 in which the cancer is detected by screening or is diagnosed as an interval cancer, where the woman has attended *r* screenings. Define C_r as the subset of the 4 screenings and 14 intervals in Figure 1, given that the number of attended screenings is equal to *r*, and add an additional category G_{rest} for cancers that are neither detected by screening nor diagnosed as interval cancer. This means that $A_{r,b} = G_{rest}$ when all screenings that could detect the cancer are false-negative and the DPCP ends after the end of the follow-up period. For given Y_l and X_J, the conditional pdf $h_{r,b}(a \mid x, y)$ of $A_{r,b}$ on domain C_r is given by:

$$h_{r,b}(a \mid x, y) = \begin{cases} q_s^b & \text{for} \quad a = G_s \quad a \in C_r \\ q_{r,i}^b & \quad\quad a = G_{r,i} \\ q_{rest}^b & \quad\quad a = G_{rest} \end{cases} \tag{4}$$

where

$$q_s^b = \begin{cases} (1-b) \cdot b^{\,s - \max\{\lceil x \rceil, 0\} - 1} & \lceil x \rceil \le t_s \le \lfloor x + y \rfloor \quad s = 1,..,r \\ 0 & \quad s = r+1,..,S \end{cases}$$

$$q_{r,i}^b = \begin{cases} b^{\,r - \max\{\lceil x \rceil, 0\}} & t_i \le \lfloor x + y \rfloor \le t_{i+1} \quad i = r,..,I \\ 0 & \quad i = 1,..,r-1 \end{cases}$$

$$q^{b}_{rest} = 1 - \sum_{s=1}^{S} q^{b}_{s} - \sum_{r=1}^{S}\sum_{i=r}^{I} q^{b}_{r,i}$$

and $\lceil d \rceil$ and $\lfloor d \rfloor$ are defined as the upper and lower entier for δ respectively.

A life history is then completely specified by the stochastic vector $Z_v = (Y_l, X, A_{r,b})$. The main interest in the model is for the unconditional pdf $p_{r,v}(a)$ of $A_{r,b}$, which can be formulated as:

$$p_{r,v}(a) = \int_{y=0}^{\infty} \int_{x=t_1-y}^{M(y)-y} J \cdot (M(y) - t_1) \cdot h_{r,b}(a \mid x, y) \cdot k(x \mid y) \cdot f_l(y) dx dy \qquad (5)$$

$$= E_{Z_v}\left\{ J \cdot (M(Y_l) - t_1) \cdot 1_{\{A_{r,b}=a\}} \right\} \quad \text{for } a \in C_r,\ r = 1,..,S$$

with the life history Z_v taken from the joint pdf $h_{r,b}(a \mid x, y) \cdot k(x \mid y) \cdot f_l(y)$. For given y, the factor $J \cdot (M(y) - t_1)$ represents the probability that the DPCP of a cancer starts in the interval of interest $[t_1 - y, M(y) - y]$. Given that a cancer occurs in this interval, the life history models the category in Figure 1 in which the cancer will be detected by screening or diagnosed as interval cancer.

Estimates of $p_{r,v}(a)$ in (5) will be obtained by simulation of a large number of disease histories Z_v using the following procedure: First a sojourn time Y_l is obtained from the sojourn time distribution $f_l(y)$. Second a moment X when a DPCP starts is obtained from $k(x \mid y)$. Finally a category $A_{r,b}$ in which the cancer is detected by screening or diagnosed as interval cancer is obtained from $h_{r,b}(a \mid x, y)$.

With the estimates for $p_{r,v}(a)$ and the total number of women $N_{r,a}$ corresponding to category a in the HIP study, where the woman has attended r screenings, the expected number of detected cancers $E_{r,a}(v)$ can be estimated by:

$$E_{r,a}(v) = N_{r,a} \cdot \frac{p_{r,v}(a)}{1 - \sum_{i \text{ before } a} p_{r,v}(i)} \qquad a \in C_r,\ r = 1,..,S \qquad (6)$$

where $p_{r,v}(i)$, i before a, indicates those categories in Figure 1 that a women has traversed before she enters category a.

2.3. Optimization of model parameters

If the microsimulation model gives an adequate representation of the natural history of breast cancer and the sensitivity of the screening test, the differences between the expected number of cancers and the observed number of cancers should be small. We use the weighted sum of squared differences between the expected and observed number of cancers as goodness of fit criterion, i.e. as a measure of the quality of the microsimulation model in representing the data recorded by the HIP trial.
In particular, for the observed numbers of cases $O_{r,a}$ in the HIP study and the corresponding expected numbers of cases $E_{r,a}(v)$ according to the microsimulation model we define the goodness of fit criterion as the statistic $c^2(v)$:

$$c^2(v) = \sum_{r=1}^{S} \sum_{a \in C_r} \frac{(O_{r,a} - E_{r,a}(v))^2}{E_{r,a}(v)} \qquad (7)$$

Note that the number of observed cases is assumed to be a realization of a Poisson distribution with as mean the corresponding expected number.

The resulting optimization problem is to find the optimal parameter vector $v^* = (l^*, b^*, J^*)$ of the microsimulation model such that the goodness of fit statistic is minimized:

$$(P_{micro}) \quad \min \; c^2(v)$$
$$\text{st } v \in V$$

where $V = \{(l, b, J) \mid l > 0, 1 \geq b \geq 0, J > 0\}$.
An optimization algorithm to find the optimal parameter vector for (P_{micro}) requires frequent evaluations of the objective function $c^2(v)$. A straightforward evaluation of $c^2(v)$ will use a sample of N individual life histories, $Z_{r,v,1}, .., Z_{r,v,N}$, simulated from pdf $g_{r,v}(z)$,

$$g_{r,v}(z) = h_{r,b}(a \mid x, y) \cdot k(x \mid y) \cdot f_l(y). \qquad (8)$$

where f_l, k and $h_{r,b}$ as defined in formulas (1), (3) and (4) respectively.

Every parameter vector *v* corresponds to a different sample $Z_{r,v,1},..,Z_{r,v,N}$ and finding the optimal parameter will therefore require many runs of the microsimulation model. Since the output of the microsimulation model is stochastic, the optimization will be hindered by noise.

Through the application of the Score Function (SF) technique [10] a standard **deterministic** optimization procedure, e.g. (quasi) Newton procedure, can be used to optimize *(P_{micro})*. The SF technique derives estimates for $p_{r,v}(a)$ in formula (5) and its derivatives for every value of the parameter vector *v*, based on a single sample, called the reference sample. This reference sample is simulated from a pdf $g_{r,v_0}(z)$ that is characterized by values of a pre-specified reference parameter vector $v_0 = (l_0, b_0, J_0)$. To estimate, for a given reference sample, $p_{r,v}(a)$ and its derivatives for other parameter vectors, the SF method does not change the sample but it changes the likelihood of the sample. This means that for a given reference sample, estimates for $p_{r,v}(a)$ and its derivatives for other parameter vectors *v* are, according to the SF method, deterministic functions of *v*.

The change in likelihood of a realization *z* of the variable vector $Z_{r,v}$ for a parameter vector *v* with respect to the reference parameter vector v_0 is modeled by the likelihood ratio:

$$W_{r,v,v_0}(z) = g_{r,v}(z) / g_{r,v_0}(z) \tag{9}$$

It is not hard to see with formula (8) and the definitions of f_l, *k* and $h_{r,b}$ in formulas (1), (3) and (4) that $W_{r,v,v_0}(z)$ is given by

$$W_{r,v,v_0}(z) = \prod_{s=1}^{r}\left(\left(\frac{1-b}{1-b_0}\right)\cdot\left(\frac{b}{b_0}\right)^{(s-\max\{\lceil x\rceil,0\}-1)\,1_{\{a=G_S\}}}\right) \prod_{i=r}^{I}\left(\left(\frac{b}{b_0}\right)^{(r-\max\{\lceil x\rceil,0\})\,1_{\{a=G_{r,i}\}}}\right)\cdot\frac{l}{l_0}\cdot\exp\left((l_0-l)\cdot y\right) \tag{10}$$

In the application for breast cancer screening we use the SF method to obtain estimates of $p_{r,v}(a)$ in formula (5) and of their derivatives with respect to $v = (l, b)$. These estimates are then used to estimate $c^2(v)$ and its derivatives with respect to *v*. Observe that the parameter *J* is not a parameter of the pdf of the life histories Z_v, but a parameter that directly influences $p_{r,v}(a)$ (formula (5)). Later on,

the (standard) SF method will be extended to include *J* and obtain estimates of $p_{r,v}(a)$ and of their derivatives with respect to $v=(l, b, J)$.

According to formula (5) $p_{r,v}(a)$ can be viewed as expectation $E\{L_{r,a}(Z_{r,v})\}$ of a certain performance measure $L_{r,a}(Z_{r,v})$

$$L_{r,a}(Z_{r,v}) = J \cdot (M(Y_l) - t_1) \cdot 1_{\{A_{r,b}=a\}} \qquad a \hat{I} C_r, r=1,\ldots,S \tag{11}$$

The SF method estimates the values of $p_{r,v}(a)$ and their derivatives for other parameters *v* using reference parameter values $v_0=(l_0, b_0)$. These reference parameters are used to obtain a reference sample $Z_{r,v_0,1},..,Z_{r,v_0,N}$ from $g_{r,v_0}(z)$.

An estimate of $p_{r,v}(a)$ is defined in the SF method by averaging the performance of the reference sample $Z_{r,v_0,1},..,Z_{r,v_0,N}$ after multiplication by the likelihood ratio:

$$\hat{p}_{r,v}(a) = \frac{1}{N}\sum_{n=1}^{N} L_{r,a}(Z_{r,v_0,n}) \cdot W_{r,v,v_0}(Z_{r,v_0,n}) \qquad a \in C_r, r=1,..,S \tag{12}$$

An estimate of the *k*-th order derivatives of $p_{r,v}(a)$ with respect to *v* is given by

$$\nabla^{(k)}\hat{p}_{r,v}(a) = \frac{1}{N}\sum_{n=1}^{N} L_{r,a}(Z_{r,v_0,n}) \cdot W_{r,v,v_0}(Z_{r,v_0,n}) \cdot S_{r,v}^{(k)}(Z_{r,v_0,n}) \tag{13}$$

for $r=1,..,S$, where $S_{r,v}^{(k)}(z)$ is called the Score Function and is given by

$$S_{r,v}^{(k)}(z) = \frac{\nabla^{(k)} g_{r,v}(z)}{g_{r,v}(z)} \tag{14}$$

Notice that $S_{r,v}^{(1)}(z)$ is a vector and $S_{r,v}^{(2)}(z)$ is a matrix.

With the pdf $g_{r,v}(z)$ as defined in (8) and by observing that

$$S^{(1)}_{r,v}(z) = \nabla \log\left(g_{r,v}(z)\right) \tag{15}$$

$$S^{(2)}_{r,v_i,v_j}(z) = \frac{\partial}{\partial v_i}\left(S^{(1)}_{r,v_j}\right) + S^{(1)}_{r,v_j} \cdot S^{(1)}_{r,v_i}$$

where $S^{(2)}_{r,v_i,v_j}(z)$ is the (*i*,*j*)-th element of $S^{(2)}_{r,v}(z)$, it is not hard to see that the score functions for this application are given by

$$S^{(1)}_{r,v}(x,y,a) = \frac{1}{l} - y \qquad a \in C_r,\ r=1,..,S \tag{16}$$

$$\frac{\partial\left(S^{(1)}_{r,l}(x,y,a)\right)}{\partial l} = -\frac{1}{l^2}$$

$$S^{(1)}_{r,b}(x,y,a) = \sum_{s=1}^{r}\left(\frac{-1}{(1-b)} + \frac{(s-\max\{\lceil x\rceil,0\}-1)}{b}\right)1_{\{a=G_S\}} + \sum_{i=r}^{I}\left(\frac{(r-\max\{\lceil x\rceil,0\})}{b}\right)1_{\{a=G_{r,i}\}}$$

$$\frac{\partial\left(S^{(1)}_{r,b}(x,y,a)\right)}{\partial b} = \sum_{s=1}^{r}\left(\frac{-1}{(1-b)^2} - \frac{(s-\max\{\lceil x\rceil,0\}-1)}{b^2}\right)1_{\{a=G_S\}} - \sum_{i=r}^{I}\left(\frac{(r-\max\{\lceil x\rceil,0\})}{b^2}\right)1_{\{a=G_{r,i}\}}$$

$$\frac{\partial\left(S^{(1)}_{r,l}(x,y,a)\right)}{\partial b} = \frac{\partial\left(S^{(1)}_{r,b}(x,y,a)\right)}{\partial l} = 0$$

Including the incidence parameter *J* in the estimations of $p_{r,v}(a)$ and their derivatives is not straightforward because *J* is not a parameter of the pdf $g_{r,v}(z)$, so the likelihood ratio and score functions with respect to *J* can not be calculated. We use two different methods to estimate the derivatives of $p_{r,v}(a)$ with respect to *J*.

First, direct differentiation of $p_{r,v}(a)$ to *J* is not difficult in this simple model. By direct differentiation of the performance measure $L_{r,a}(Z_{r,v})$ defined in (11), estimates for $\frac{\partial^{(k)} p_{r,v}(a)}{(\partial J)^{(k)}}$ can be calculated as follows for *k*=1,2:

$$\frac{\partial \hat{p}_{r,v}(a)}{\partial J} = \frac{1}{N}\sum_{n=1}^{N}\left(M(Y_{l_0,n}) - t_1\right)\cdot 1_{\{A_{r,b,n}=a\}} \qquad a \in C_r,\ r=1,..,S \tag{17}$$

$$\frac{\partial^{(2)} \hat{p}_{r,v}(a)}{(\partial J)^2} = 0$$

Alternatively, J can be estimated using the 'push-out' technique. The 'push-out' technique implies that a parameter that influences the performance measures directly (instead of indirectly through the underlying pdf) is pushed out of the performance measures into a pdf, so that the standard SF method can be applied to estimate this parameter. We used this technique for J, by introducing an extra variable Q_q for the incidence, i.e. the constant J in the performance measure $L_{r,a}(Z_{r,v})$ in formula (11) is replaced by Q_q. The optimal value of J will be estimated as Q_q's mean.

We assume a Weibull distribution for Q_q with scale parameter q and a fixed shape parameter a.. The pdf of Q_q is given by

$$m_q(q) = \frac{a}{q} \cdot \left(\frac{q}{q}\right)^{a-1} \exp\left(-\left(\frac{q}{q}\right)^{a}\right). \tag{18}$$

Estimates of $p_{r,v}(a)$ and derivatives with respect to q can be calculated using the standard SF method. J and q are directly related through the relation $J = q \cdot c$, where $c=\Gamma\left(1+\frac{1}{a}\right)$ is constant for fixed a and $\Gamma(t)=\int_0^\infty z^{t-1}\exp(-tz)dz$ is the Gamma function.

In the 'push-out' variant the parameter vector is $v = (l, b, q)$ and the vector of stochastic variables is $Z_{r,v} = (Y_l, X, A_{r,b}, Q_q)$ with pdf $g_{r,v}(z)$, where r is the number of attended screenings, as defined as follows:

$$g_{r,v}(z) = m_q(q) \cdot h_{r,b}(a \mid b, y) \cdot k(b \mid y) \cdot f_l(y) \tag{19}$$

Estimates of $p_{r,v}(a)$ and their derivatives can be calculated using the SF method, using formula (5), with $Z_{r,v_0,1},..,Z_{r,v_0,N}$ from $g_{r,v_0}(z)$ as defined in formula (19), the constant J in the performance measures $L_{r,a}(Z_{r,v})$ in formula (11) replaced by Q_q, and the likelihood ratio $W_{r,v,v0}(z)$ as defined in formula (20). The score function $S_{r,v}^{(k)}(z), k=1,2$ can be calculated using formulas (15), (16) and (21).

$$W_{r,v,v_0}(z) = \prod_{s=1}^{r}\left(\left(\frac{1-b}{1-b_0}\right)\cdot\left(\frac{b}{b_0}\right)^{(s-\max\{\lceil x\rceil,0\}-1)\,1_{\{a=G_s\}}}\right)\cdot\prod_{i=r}^{I}\left(\left(\frac{b}{b_0}\right)^{(r-\max\{\lceil x\rceil,0\})\,1_{\{a=G_{r,i}\}}}\right)\cdot$$
$$\frac{l}{l_0}\cdot\exp((l_0-l)\cdot y)\cdot\left(\frac{q_0}{q}\right)^{a}\cdot\exp\left(\left(\frac{q}{q_0}\right)^{a}-\left(\frac{q}{q}\right)^{a}\right) \tag{20}$$

$$S_{r,q}^{(1)}(z) = -\left(\frac{a}{q}\right)+\left(\frac{a}{q}\right)\cdot\left(\frac{q}{q}\right)^{a} \tag{21}$$

$$\frac{\partial\left(S^{(1)}_{r,q}(z)\right)}{\partial q} = \left(\frac{a}{q^2}\right) - \left(\frac{a}{q}\right)\cdot\left(\frac{a+1}{q}\right)\cdot\left(\frac{q}{q}\right)^a$$

$$\frac{\partial\left(S^{(1)}_{r,v_i}(z)\right)}{\partial v_j} = 0 \qquad i \neq j, i \leq 3, j \leq 3$$

3. Tests of the score function method for microsimulation

The one stage breast cancer model as proposed by Walter and Day for the HIP data set can be solved analytically. Using the analytical model we find the optimal parameter values $v^* = (l^*, b^*, J^*) =$ (0.614, 0.129, 0.00213) and the goodness of fit value $c^2(v^*) = 13.343$.

The optimal parameters found using the microsimulation model and the score function technique are compared to these analytical optimal values. The non-linear deterministic optimization problem that remains when the SF technique is applied to the generated reference sample of life histories is solved using the standard Quasi-Newton algorithm [14].

The precision of a solution v_{SF} resulting from the optimization using the SF method is measured by the error of the goodness of fit of this solution compared to the goodness of fit of the analytical optimum:

$$e = c^2(v_{SF}) - c^2(v^*)$$

where the goodness of fit for both v^* and v_{SF} is evaluated using the analytical model.

There are 5 models in our analysis:

1. Only the length Y_l of the DPCP is simulated, optimization of parameter l
2. Simulation of life histories $Z_{l,b,r} = (Y_l, X, A_{r,b})$, where r, r=1,..,4, is the number of screenings attended, optimization of parameter l.
3. Simulation of life histories $Z_{l,b,r} = (Y_l, X, A_{r,b})$, optimization of parameters l and b.
4. Simulation of life histories $Z_{l,b,r} = (Y_l, X, A_{r,b})$, optimization of parameters l, b and J.
5. Simulation of life histories $Z_{l,b,r} = (Y_l, X, A_{r,b})$, simulation of stochastic variable Q_J, representing the incidence, from a Weibull distribution with fixed shape parameter α=3, optimization of parameters l, b and J.

The first model arises from the analytical model, where b and J are determined analytically. Only the length of the DPCP is simulated. In the second model, the complete life histories $Z_{l,b,r}$ are simulated,

but b and J are assumed to be known. Thus, the life histories $Z_{l,b,r}$ are generated with the optimal parameter values b^*=0.129 and J^*=0.00213. The optimal value of l is determined with the SF – Quasi Newton technique. In model 3 (4) the parameter value b (b and J) is also included in the optimization. Finally, in model 5 is the 'push-out' implementation, where the parameter J is replaced by the stochastic variable Q_J with a Weibull distribution with fixed shape parameter a =3.

The SF technique includes the choice of a reference sample. This sample is related to the reference parameter values v_0= (l_0, b_0, J_0). We compare the performance of the models for two reference parameters. The first reference parameters are close to the optimal solution and are set to v_0=(l_0, b_0, J_0)=(0.6, 0.1, 0.0025), and the second reference parameters are located farther from the optimum, v_0=(l_0, b_0, J_0)=(0.4, 0.4, 0.003). We test how the precision and the computational effort depend on the value of the reference parameters.

To test the impact of the simulation size on precision and computational effort, the parameters are estimated with simulation size equal to N =10,000 and N =100,000 life histories. The simulation is repeated 20 times for each combination of reference parameters and simulation size and the average results are recorded together with the standard deviation. The computation time is measured in average number of seconds needed before the Quasi-Newton algorithm terminates. The tests are performed on a stand alone Pentium 3 PC, 550 MHz with 256 MB internal memory. Table ** a shows the results for the five models, with 10,000 simulated life histories. The columns show the reference parameters v_0=(l_0, b_0, J_0), the average value of the optimal parameters v_{SF} =(l_{SF}, b_{SF}, J_{SF}) resulting from the SF-Quasi Newton method, the average error, and the average running times (in seconds). The averages are taken over the 20 simulation runs. Table 1b shows the same results for 100,000 simulated life histories.

Table 1a Simulation size 10,000

Reference parameters	Average SF solution (st. dev.)	Average error (st. dev.)	Average running time (st. dev.)
l_0 =0.6 (model 1)	0.614 (0.005)	0.007 (0.007)	3 (1)
l_0 =0.6 (model 2)	(0.614) (0.008)	0.016 (0.017)	3 (1)
(l_0, b_0)=(0.6, 0.1) (model 3)	(0.628, 0.113) (0.026, 0.033)	0.076 (0.083)	4 (1)
(l_0, b_0, J_0)=(0.6, 0.1, 0.0025) (model 4)	(0.622, 0.119, 0.00213) (0.039, 0.044, 0.00002)	0.092 (0.134)	6 (1)
(l_0, b_0, J_0)=(0.6, 0.1, 0.0025) (model 5)	(0.620, 0.127, 0.00213) (0.044, 0.051, 0.00002)	0,129 (0,145)	16 (1)
l_0 =0.4 (model 1)	0.614 (0.005)	0.008 (0.011)	6 (1)
l_0 =0.4 (model 2)	(0.614) (0.008)	0.015 (0.017)	5 (1)
(l_0, b_0)=(0.4, 0.4) (model 3)	(0.607, 0.135) (0.026, 0.024)	0.055 (0.039)	6 (1)
(l_0, b_0, J_0)=(0.4, 0.4, 0.003) (model 4)	(0.610, 0.135, 0.00213) (0.034, 0.035, 0.00002)	0.102 (0.091)	10 (1)
(l_0, b_0, J_0)=(0.4, 0.4, 0.003) (model 5)	(0.630, 0.117, 0.00214) (0.039, 0.041, 0.00002)	0.124 (0.093)	26 (3)

Table 1bSimulation size 100,000

Reference parameters	Average SF solution (st. dev.)	Average error (st. dev.)	Average running time (st. dev.)
l_0 =0.6 (model 1)	0.615 (0.002)	0.001 (0.001)	34 (7)
l_0 =0.6 (model 2)	0.613 (0.003)	0.002 (0.003)	26 (5)
(l_0, b_0)=(0.6, 0.1) (model 3)	(0.614, 0.130) (0.007, 0.009)	0.006 (0.005)	38 (7)
(l_0, b_0, J_0)=(0.6, 0.1, 0.0025) (model 4)	(0.608, 0.136, 0.00213) (0.012, 0.014, 0.00001)	0.012 (0.017)	65 (3)
(l_0, b_0, J_0)=(0.6, 0.1, 0.0025) (model 5)	(0.613, 0.130, 0.00213) (0.013, 0.013, 0.00001)	0.008 (0.010)	158 (14)
l_0 =0.4 (model 1)	0.613 (0.002)	0.001 (0.002)	61 (6)
l_0 =0.4 (model 2)	0.613 (0.003)	0.002 (0.003)	53 (7)
(l_0, b_0)=(0.4, 0.4) (model 3)	(0.614, 0.129) (0.009, 0.009)	0.007 (0.005)	62 (7)
(l_0, b_0, J_0)=(0.4, 0.4, 0.003) (model 4)	(0.612, 0.131, 0.00213) (0.010, 0.011, 0.00000)	0.008 (0.007)	112 (8)
(l_0, b_0, J_0)=(0.4, 0.4, 0.003) (model 5)	(0.614, 0.130, 0.00213) (0.010, 0.010, 0.00001)	0.012 (0.007)	245 (47)

The results clearly show that the score function technique combined with the Quasi-Newton method can optimize the model parameters with very small error. Including more parameters in the optimization decreases the precision of the parameter estimates and increases the average

computational time. However, the estimates stay within close range of the optimal values. The method never finds a value far away from the optimum, both for the reference parameter close to the optimum and for the reference parameter further from the optimal value. The computational times increase when more parameters are included in the optimization, and including an extra stochastic process in the simulation slows the optimization method down significantly.

Parameter *J* is very close to its optimal value for all models and every test setup. Notice the different performance between model 4 where *J* is estimated directly and model 5 where *J* is estimated through the parameter of the push-out distribution. The computational time increases significantly but the estimate of the parameter *J* is equally good for both models. The optimization of parameter l is strongly influenced by the model under consideration, especially for the simulation size 10,000. The variance in the parameter estimate grows with the complexity of the model, and a larger simulation size is preferred for parameter l. The parameter estimates for *b* also benefit from a larger simulation size. All parameter estimates seem unaffected by the chosen reference parameter with respect to precision of the estimates. However the computational effort is higher for the reference parameters further from the optimum. A tenfold increase in the simulation size will lead to a similar increase in computational time but also to an error that is tenfold smaller. These results suggest that both relationships are linear in the size of the simulation.

In our analysis we use a Quasi-Newton algorithm to estimate the parameters. The Quasi-Newton algorithm is a search method that determines a search direction based on the gradient and an approximation of the Hessian, and uses a line search to determine the stepsize in the search direction. The Newton algorithm uses the gradient and the Hessian to determine an alternative for the current point in the feasible region. The Newton algorithm therefore also needs the Hessian of the objective function. Although the Hessian is not given for our application in an analytical form, the SF technique can also provide an estimate of the Hessian based on the reference parameters. To test whether the precision or computational effort could be improved by using the Newton algorithm instead of the Quasi-Newton algorithm, we estimated parameters (l, *b*, *J)* for model 5 on the basis of 20 optimization runs for both reference values and simulation size equal to *N*=100,000 life histories. For reference values (l_0, b_0, J_0)=(0.6, 0.1, 0.0025), i.e. the values close to the optimum, the Newton algorithm performs as good as the Quasi-Newton algorithm regarding the precision, but is twice as fast. But for the reference parameters (l_0, b_0, J_0)=(0.4, 0.4, 0.003) located further from the optimum the Newton algorithm can not find a solution in 12 of the 20 optimization runs. The Newton method is thus very sensitive to the choice of the reference parameter and seems to be suitable only if there is some information on the approximate optimal value of the parameters.

We also considered the influence of the parameters of the Weibull distribution on the performance of the optimization procedure. In model 5 the incidence parameter J is replaced by a stochastic variable Q_J that has a Weibull distribution with fixed shape parameter α=3. To test the impact of this choice for α on the solution, we estimate (l,b,J) for model 5 with α=1, which means that the Weibull distribution reduces to an exponential distribution. The shape of the exponential distribution will result in a higher probability to sample smaller and larger values of J compared to the Weibull distribution with shape parameter α=3 that has a higher probability to sample values close to the average. On the basis of 20 optimization runs for both reference values and a simulation size equal to N=100,000 life histories we tested the sensitivity of the performance to the choice of α. For the reference parameters (l_0, b_0, J_0)=(0.6, 0.1, 0.0025) close to the optimum the average precision was worse (0.014) for α=1 than for α=3, but for the reference values (l_0, b_0, J_0)=(0.4, 0.4, 0.003) located further from the optimum the average precision was better (0.09). For both reference values the computational effort was similar to those for α=3, but we must conclude that the shape parameter can have some influence on the precision. Without any knowledge on the values of the model parameters that one wants to estimate a low value of the shape parameter α seems to be preferable.

4. Discussion

The Score Function technique in combination with the Quasi-Newton algorithm appears to be an efficient method for fitting a simple microsimulation cancer screening model to data. The results show that the precision of the derived optimal value can be improved by taking a larger sample of simulated life histories: the error in the performance (goodness of fit) function is approximately 10 times smaller if the sample is 10 times larger. For a given sample size, the precision will decrease when more parameters are fitted simultaneously. The simple model considered can completely be modeled analytically or can be modeled partly through stochastic processes. When the simulation includes more stochastic processes the precision of the parameter estimates also decreases. A special case is the application of the 'push-out' technique, which implies that a parameter that influences the performance measures directly (instead of indirectly through the underlying pdf) is pushed out of the performance measures into a convenient pdf, so that the standard SF method can be applied to estimate such a parameter. The incidence parameter J is such a parameter in the cancer-screening model. This parameter has a relatively simple relation with the goodness of fit measure, and we therefore used direct estimates of gradients with respect to J in the optimization procedure in model 4. In model 5, we explored use of the push-out technique for this parameter J, and found that the precision of the parameter estimates were similar, but that the computational effort is about two times higher for the push-out technique.

The Score Function method comes with a nice bonus: it will produce an (approximate) confidence area for the model parameters at only little extra expense. Using the SF method a second order approximation for $c^2(l,b)$ is estimated in the analytical optimum (l=0.614, b=0.129) and this approximation is used to estimate the 95% confidence region for (l,b), see Figure 2. The figure shows that the HIP data still leave room for uncertainty: a proportion of up to $b=0.45$ of pre-clinical breast cancers could be missed by the screening tests used around 1970, and the mean duration of the screen-detectable phase (DPCP) can be between 1.17 years (l=0.858) and 3.17 years (l=0.315). Note that the quality of mammography has improved considerably since then: much smaller tumors can now be detected, resulting in a longer (average) duration of the DPCP, and the sensitivity of the screening test has increased.

==Insert Figure 2==

The same breast cancer microsimulation model has been used to investigate the performance of different versions of the Nelder and Mead Simplex method (NMSM) [6]. Model 4 of the SF-Quasi Newton technique for a reference sample of 100,000 has a mean error that is 100x smaller than the mean error of the original setting of the NMSM (benchmark version). The benchmark NMSM will on average evaluate 15.10^5 function values during an optimization run; the reference sample of 100,000 life histories used in the SF method led to an average of 36.10^5 function evaluations during an optimization run. An adapted version of the NMSM, in which the number of life histories per simulation run increases during the optimization, has a smaller average error but needs a total of 50.10^5 function evaluations. Still the error is 25x larger than the mean error based on the SF method. Clearly, the SF method gives a more precise estimate of the parameters using approximately the same computational effort.

The simple optimization problem of fitting the one-stage breast cancer model to the HIP trial data was chosen as test problem for exploring the feasibility and efficiency of the SF approach for cancer screening models for two reasons: the model and the data are typical for cancer screening, and the true optimum can be derived from an analytical version of the model. The implementation of this analytical model to the HIP data is not too difficult, but when extensions to the model are required the implementation will rapidly become complicated, and microsimulation will be a more feasible approach. The effort for constructing an efficient microsimulation version is relatively large for this simple one-stage breast cancer model. The incidence hazard (J) can in principle give rise to a long-lasting DPCP that starts many years before a first screening test. This problem was resolved by reversing the order in which

model variables are generated: first generate the duration of the DPCP, then generate when this DPCP will start.

A typical extension to the simple model is to add a dimension to model the age of the woman. Several characteristics of breast cancer are age-dependent: the incidence increases with age, the (mean) duration of the DPCP increases with age, the stage distribution of clinically diagnosed cancers, and the survival probability of breast cancer patients vary with age. Similar age-specific screening models can be developed for cancer of the cervix, colon, prostate and lung [2]. Such models can e.g. address the important question of finding the optimal ages for screening [15],[16]. In age-specific models, there is no need to reverse the simulation procedure: the age at which DPCP starts has a natural lower limit (age 0). The age for the start of the DPCP can be generated from a postulated age-specific incidence hazard function, and then the duration of the DPCP can be generated, possibly from an age-dependent sojourn time distribution. Hence, the microsimulation procedure is more straightforward for an age-dependent one-stage breast cancer than for the simpler age-independent one-stage model considered in this paper. Implementation of the SF method to the age-dependent breast cancer model is expected to be straightforward. Implementation of further refinements to the model are also expected to be relatively easy, but will of course lead to further reductions in the precision of the estimated optimum. Larger sample runs can compensate for the loss in precision at the cost of increased computational effort.
The SF method is now being used in developing and testing a more detailed breast-cancer-screening model.

Our results show that the SF method is useful in optimizing parameters of pdfs that govern the simulation of life histories including the impact of screening. An initial sample of life histories is generated which is used throughout the optimization procedure. For each new set of parameter values, a likelihood ratio is applied to adapt the contribution of each history to the overall performance measure (in our case the goodness of fit), and the score function will produce the corresponding gradients and, if desired, the Hessian matrix. The random nature of the sample will result in deviations from the true optimum value, which can be decreased by taking a larger sample. Once the sample is generated, the optimization problem is no longer stochastic and can be solved numerically by using derivative-based optimization methods for deterministic problems that use the gradients and Hessians derived with the SF method.
Other perturbation methods such as the quasi-gradient approach will derive estimates of the gradient by investigating how a small change in the parameter value will change the generation of life histories and thus affect the performance measure. This can be combined with a Stochastic Approximation type of optimization routine (using a small number of simulated histories per optimization step) to find the optimum. A condition for applying this method is that perturbations in the parameter estimates should not lead to sudden large changes in the performance measure, i.e. no discontinuities should occur in the

contribution of a history to the performance function [17]. This approach has been tested for a relatively simple model for cervical cancer screening with good results [18]. However, meeting the continuity condition required adaptations to the simulation procedure that are specific for the model used; each extension to the model would require a complete re-investigation of the model.

The SF method is expected to be very useful for other medical decision models than the breast cancer-screening model presented here. Models for clinical decision making typically consider life histories where symptoms and complaints occur at a certain point in time. A decision has to be made regarding a diagnostic strategy or a therapy has to be chosen. Increasingly, microsimulation is being used in evaluating these models [19], again by generating large numbers of individual clinical histories, and the SF method could be explored as a method for finding the optimal policy.

A condition for using the using the SF method for microsimulation medical decision making models is that life histories are mutually independent. This is not the case for infectious disease models such as ONCHOSIM [20] and LYMFASIM [21]. For these models we will still have to rely on direct search methods such as the NMSM approach or the Response Surface methodology [21].

References

[1] G.J. van Oortmarssen, J.D.F. Habbema, J.Th.N. Lubbe and P.J. van der Maas, A model-based analysis of the HIP project for breast cancer screening, Int. J. Cancer. 46 (1990) 207-213

[2] F. Loeve, R. Boer, G.J. van Oortmarssen, M. van Ballegooijen and J.D.F. Habbema, The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening, Comp. Biomed. Res. 32 (1999) 13-33

[3] N.E. Day and S.D. Walter, Simplified models of screening for chronic disease: estimation procedures from mass screening programmes, Biometrics 40 (1984) 1-14

[4] M.W. Trosset, I Know It When I See It: Toward a Definition of Direct Search Methods, SIAG/OPT Views-and-News 9 (1997)

[5] J.A. Nelder and R. Mead, A simplex method for function minimization, The Computer Journal 7 (1965) 308-313

[6] H.G. Neddermeijer, G.J. van Oortmarssen, N. Piersma, R. Dekker and J.D.F. Habbema, Adaptive extensions of the Nelder and Mead Simplex Method for optimization of stochastic simulation models, Econometric Institute Report EI2000-22/A, faculty of Economics, Erasmus University Rotterdam, The Netherlands (2000).

[7] R.H. Myers and D.C. Montgomery, *Response surface methodology: process and product optimization using designed experiments*, New York: John Wiley & Sons, 1995

[8] J.P.C. Kleijnen, Experimental design for sensitivity analysis, optimization and validation of simulation models, in: *Handbook of Simulation: principles, methodology, advances, applications and practice*, ed. J. Banks, 1998, New York: John Wiley & Sons, pp. 173-223.

[9] J.C. Spall, A stochastic approximation technique for generating maximum likelihood parameter estimates, in: *Proceedings of the American Control Conference*, 1161-1167.

[10] R.Y. Rubinstein and A.Shapiro, *Discrete event systems. Sensitivity analysis and stochastic optimization by the score function method*, New York: John Wiley & Sons, 1993

[11] J.P.C. Kleijnen and R.Y. Rubenstein, Optimization and sensitivity analysis of computer simulation models by the score function method, Eur. J. of Operations Research 88 (1996) 413-427

[12] S.D. Walter and N.E. Day, Estimation of the duration of a preclinical stage using screening data, Am. J. of Epidemiol. 118 (1983) 865-886

[13] S. Shapiro, J.D. Goldberg and G.B. Hutchinson, Lead time in breast cancer detection and implications for periodicity of screening, Am. J. of Epidemiol. 100 (1974) 357-366

[14] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical recipes in C: the art of scientific computing*, Cambridge University Press, 1992

[15] R. Boer, H.J. de Koning, G.J. van Oortmarssen and P.J. van der Maas, In search of the best upper age limit for breast cancer screening, Eur. J. Cancer 31A (1995) 2040-2043

[16] Boer, H.J. de Koning and P.J. van der Maas, A longer breast carcinoma screening interval for women age older than 65 years?, Cancer 86 (1999) 1506-1510

[17] P. L'Ecuyer, An overview of derivative estimation, in: *Proceedings of the 1991 Winter Simulation Conference*, eds. B.L. Nelson, W.D. Kelton and G.M. Clark, 1991, pp. 207-217.

[18] G.J. van Oortmarssen, Y.M. Ermoliev, Stochastic Optimisation of Screening Strategies for Preventing Irreversible Changes, IIASA Working Paper WP-94-124, IIASA, Laxenburg, (1994).

[19] A.G. Chessa, R. Dekker, B. van Vliet, E.W. Steyerberg, J.D. Habbema, Correlations in uncertainty analysis for medical decision making: an application to heart-valve replacement, Med. Decis. Making. 19 (1999) 276-86

[20] A.P. Plaisier, G.J. van Oortmarssen, J.D.F. Habbema, J. Remme and E.S. Alley, ONCHOSIM: a model and computer simulation program for the transmission and control of onchocerciasis, Comput. Methods Programs Biomed. 31 (1990) 43-56

[21] A.P. Plaisier, S. Subramanian, P.K. Das, W. Souza, T. Lapa, A.F. Furtado, C.P.B. van der Ploeg, J.D.F. Habbema and G.J. van Oortmarssen, The LYMFASIM simulation program for modelling lymphatic filariasis and its control, Meth. Inform. Med. 37 (1998) 97-108

[22] H.G. Neddermeijer, G.J. van Oortmarssen, N. Piersma and R. Dekker, A framework for response surface methodology for simulation optimization, Econometric Institute Report EI2000-14/A, Faculty of Economics, Erasmus University Rotterdam, The Netherlands (2000).

Figure 1: Data for the first five years of the HIP trial of breast cancer screening. There are 4 screenings and 14 follow up interval groups (S=4, I=5). For each screening we list the detection rates (per 1000 women examined) and between brackets the number of cancers detected in the HIP trial. For each interval group we list the incidence (per 1000 women-years) and between brackets the number of cancers diagnosed in the HIP trial.

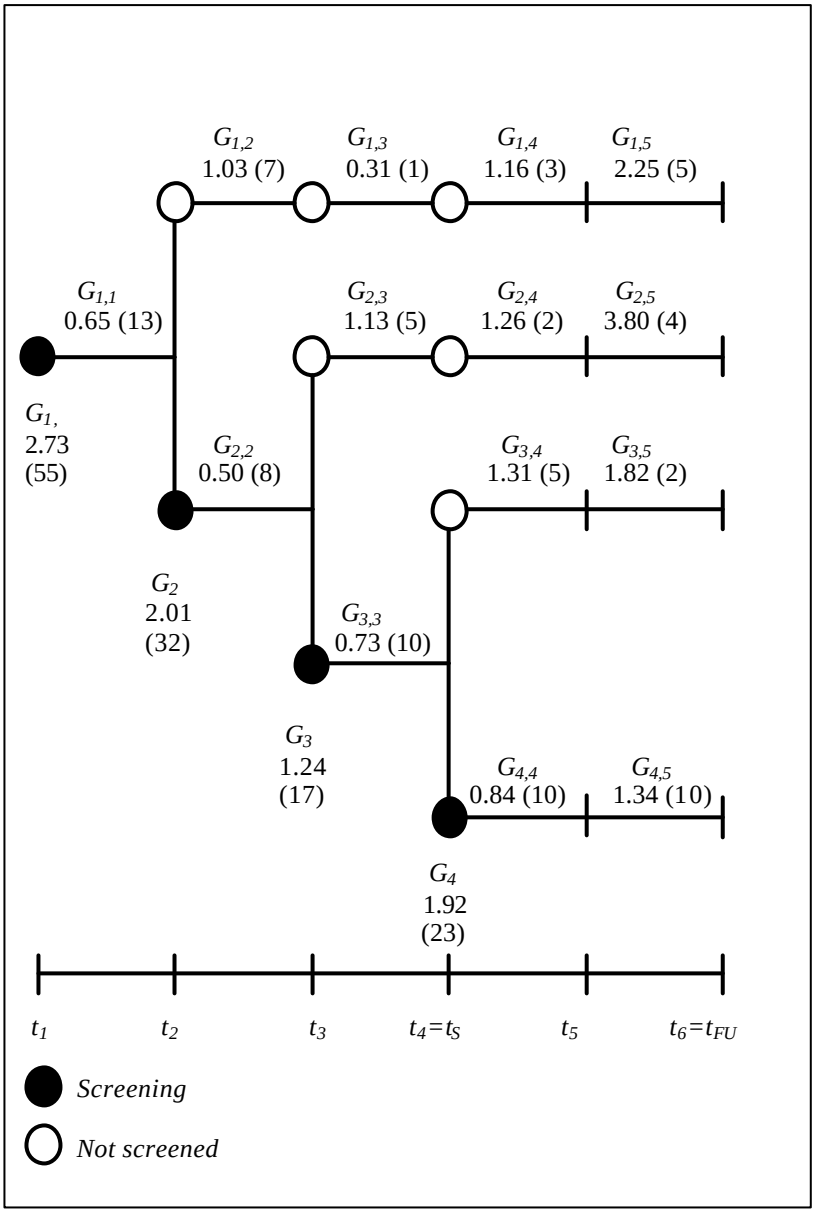

Figure 2: estimated 95% confidence region for (*l*,*b*)