STATISTICS IN MEDICINE
Statist. Med. **17**, 2509–2523 (1998)

ASSESSING UNCERTAINTY IN MICROSIMULATION MODELLING WITH APPLICATION TO CANCER SCREENING INTERVENTIONS

KATHLEEN A. CRONIN[1]*, JULIE M. LEGLER[2] AND RUTH D. ETZIONI[3]

[1]*Biometry Branch, National Cancer Institute, EPN 313, 6130 Executive Plaza North, Bethesda, MD 20892, U.S.A.*
[2]*Applied Research Branch, National Cancer Institute, EPN 313, 6130 Executive Plaza North, Bethesda, MD 20892, U.S.A.*
[3]*Fred Hutchinson Cancer Research Center, 1124 Columbia Street, MP-665, Seattle, WA 98104-2092, U.S.A.*

SUMMARY

Microsimulation is fast becoming the approach of choice for modelling and analysing complex processes in the absence of mathematical tractability. While this approach has been developed and promoted in engineering contexts for some time, it has more recently found a place in the mainstream of the study of chronic disease interventions such as cancer screening. The construction of a simulation model requires the specification of a model structure and sets of parameter values, both of which may have a considerable amount of uncertainty associated with them. This uncertainty is rarely quantified when reporting microsimulation results. We suggest a Bayesian approach and assume a parametric probability distribution to mathematically express the uncertainty related to model parameters. First, we design a simulation experiment to achieve good coverage of the parameter space. Second, we model a response surface for the outcome of interest as a function of the model parameters using the simulation results. Third, we summarize the variability in the outcome of interest, including variation due to parameter uncertainty, using the response surface in combination with parameter probability distributions. We illustrate the proposed method with an application of a microsimulator designed to investigate the effect of prostate specific antigen (PSA) screening on prostate cancer mortality rates. © 1998 John Wiley & Sons, Ltd.

1. INTRODUCTION

Microsimulation models are becoming an indispensable tool for understanding the natural history of disease and predicting the impact of complex medical interventions. The microsimulation approach provides an alternative to traditional analytic methods,[1–3] which may be intractable when interventions are administered with a complex schedule. Some examples of microsimulation approaches applied to cancer research include: modelling of the natural history of disease and the introduction of a screening intervention for breast, cervical, colorectal and prostate cancer;[4–7] identifying an efficient protocol for ovarian cancer screening;[8] and analysing the costs and benefits of introducing a mammography screening programme.[9]

* Correspondence to: Kathleen A. Cronin, Biometry Branch, National Cancer Institute, EPN 313, 6130 Executive Plaza North, Bethesda, MD 20892, U.S.A. E-mail: cronink@dcpcepn.nci.nih.gov

Contract/grant sponsor: NCI
Contract/grant number: N01-CN-05230

CCC 0277–6715/98/212509–15$17.50

Received July 1997
Accepted February 1998

With microsimulation, we simulate a large number of individual life histories using random numbers and we summarize the resulting quantities of interest using a count, proportion or some other function of the simulated data. There are two major goals of microsimulation in cancer incidence and mortality trend analyses. The first is *model fitting*. Often one considers data from the years before the advent of screening and uses microsimulation to construct reasonable models for the cancer's natural history uncontaminated by screening effects.[5] Repeated runs with different model parameter values may suggest a set of plausible parameter values that reasonably explain the observed data.

A second goal of microsimulation is *experimentation*. One can use the simulation model to predict *changes* as a result of an intervention. For example, one may have interest in predicting the impact of a newly proposed cancer screening test on a population. The outcome of interest may be lifeyears saved or number of deaths prevented. This is accomplished by first assuming a model structure, distribution form, and a set of specified values for the model parameters. Then one runs the model with and without the intervention applied to simulated life histories. Unlike the first goal described above, where the inference focuses on the model parameters, here the inference concerns the estimation of the change in one or more critical characteristics of the model results.

Simulation results are subject to three sources of variation.[10,11] The first is stochastic variation. A stochastic simulation employs a random number generator so that individual runs lead to different outcomes. This reflects the inherent heterogeneity in a population. A second source of variation in microsimulation is associated with the fact that model parameters are estimated, not known. Assessing the effects of this parameter uncertainty is the focus of this paper. A third important source of variation, not addressed in this paper, is associated with the choice of the specified model structure.

Parameter uncertainty can be of particular concern in epidemiologic simulations[12] because, unlike engineering applications, with medical applications it is quite difficult to perform experiments that provide further knowledge concerning model parameters. Typically, a sensitivity analysis is performed by varying values for the parameters one at a time and noting the variation in simulated data from individual microsimulation runs. Often it is the only way in which parameter uncertainty is reflected in an analysis.

In this paper we take sensitivity analysis several steps further. Our objective is to provide an analytical framework to assess the variation in data produced by a microsimulation that explicitly incorporates parameter uncertainty. To achieve this, we propose a method that provides a useful summarization and illustration of sensitivity analysis results, and also produces an estimate and standard error for the intervention effect of interest that incorporates parameter uncertainty. We use a Bayesian approach and treat model parameters as random variables. We assume a probability distribution for the model parameters to represent parameter uncertainty. The probability distribution can be based on either a combination of expert opinion and prior belief or observed data (for example, we may derive it from a set of plausible parameter values determined by fitting a model using data as described above for the first goal of microsimulation). There is usually not a closed form for the density function for the simulated effects, hence we have the need for a computational approach.

Our approach is a logical synthesis of ideas that have appeared in the methodological literature on model uncertainty in epidemiological research. Ackerman[12–15] has published a four part tutorial on micropopulation simulations. This series covers both basic information on simulation and issues specific to micropopulations and epidemiologic research. Ma *et al.*[16] and Ma and

Ackerman[17] discussed parameter sensitivity analysis and the use of a Latin squares hypercube sampling design. Doubilet *et al.*[18] suggested a probabilistic sensitivity analysis using Monte Carlo simulation applied to a decision tree. Each probability and utility of the decision tree was represented by a probability distribution rather than a single value.

The general problem of model uncertainty has been addressed in a variety of fields. McKay *et al.*[10,11] examined the source of simulation uncertainty in some detail and estimated the contribution of each source (stochastic, parameter and model structure) to the overall prediction uncertainty. Faraway[19] looked at bootstrap, jack-knife and sample splitting approaches to adjust for overoptimistic confidence in model and parameter estimates due to model selection based on data analysis. Draper[20] and Madigan *et al.*[21] addressed model structure uncertainty by suggesting Bayesian techniques that place probabilities on possible model structures to obtain a single variance estimate. One can easily apply these methods when the form of the output variable density function is known after conditioning on model structure and parameter values. Application areas include economics and cost effectiveness analysis,[22] policy analysis[23] and system safety.[24,25]

We describe a method for quantifying the impact of parameter uncertainty on simulated data. We compare the performance of the proposed method with a parametric bootstrap procedure using a hypothetical example of a breast cancer screening intervention. Finally, we apply our method to an application that addresses a critical question in cancer research, the effect of PSA screening and its effect on prostate cancer mortality trends.[26]

2. METHODS

We construct a microsimulator model by specifying values for the model parameters and generating life histories for the simulated individuals by drawing random numbers. Model parameters may include mean pre-clinical time, test sensitivity, and benefit from screening in the form of a hazard ratio. For each run, the microsimulator produces simulated data such as the number of cases detected by screening, number of deaths averted by screening, or some other quantity of interest, depending on the application.

Often, the initial step in a microsimulation is to use maximum likelihood methods to fit the model and to obtain point estimates for unknown model parameters. Next, we consider these model parameter values fixed and we simulate and compare data for different intervention strategies. Multiple runs using a fixed set of MLEs reflects only the stochastic variation associated with the simulation model, not the uncertainty related to the parameter estimates.

One method of incorporating the uncertainty regarding the parameters into the analysis is a parametric bootstrap procedure.[27] This procedure generates sets of parameter values from a multivariate distribution and runs the simulation program using each set of generated model parameters. A bootstrap analysis yields estimates for variances and confidence intervals for the simulated data characteristics of interest that include the effects of parameter uncertainty. One advantage of this approach is that there is no need to assume a model for the relationship between simulation outcomes and parameter values. A disadvantage is that many runs may be needed to obtain an accurate estimate for the variance and confidence interval of the outcome of interest. Thus the implementation of this method may be cumbersome due to the number of simulation runs required.

Another method used in sensitivity analysis is the experimental design approach of Latin squares sampling.[12,28] McKay *et al.*[28] suggests a Latin hypercube sampling method where each dimension of the hypercube is divided into strata with equal marginal probability. Ackerman[12]

suggests sampling points within a hypercube representing possible values of the model parameters and then fitting a regression to the mean output at the simulated points to give a response surface. This response surface reflects the change in the mean simulated data as a function of the model parameters.

We suggest an experimental design procedure that produces a single estimate for mean and variance that includes parameter uncertainty but requires a more manageable number of simulation runs than bootstrapping. First, we use Latin squares sampling to choose points from the parameter space; however, one may apply other sampling methods as well. Then we run the simulation multiple times at each of the points chosen and we fit a response surface to the simulation outcomes using regression methods. Finally, we use this fitted surface, along with the probability distribution for the model parameters, to estimate the mean and the variance of the simulated outcome of interest. We use a grid that is equally spaced over the parameter space, rather than divide the parameter space into regions of equal probability. This was done to separate the fitting of the response surface from the probability distribution for the model parameters. The advantages of this are examined in Section 5.

The five steps listed below describe the procedure in detail. Let y equal the simulation point estimate of interest, Λ represent the vector of model parameters, and $f(\Lambda)$ the multivariate probability distribution of Λ:

Step 1. Sample L points in the parameter space using a Latin hypercube sampling design.

Step 2. Run the simulation M times at each sampled point.

Step 3. Fit a response hypersurface to the LM simulated point estimates. This approximates the surface $g(\Lambda)$ where $E(y|\Lambda) = g(\Lambda)$.

Step 4. Integrate the estimated response surface, $\hat{g}(\Lambda)$, with respect to the multivariate probability distribution of the model parameters, $f(\Lambda)$, to obtain an estimate of the overall mean and variance for the response surface with respect to the probability distribution for Λ using:

$$E_\Lambda(E(y|\Lambda)) = \int_\Lambda g(\Lambda) f(\Lambda) \mathrm{d}\Lambda$$

$$\mathrm{var}_\Lambda(E(y|\Lambda)) = \int_\Lambda g(\Lambda)^2 f(\Lambda) \mathrm{d}\Lambda - E_\Lambda(E(y|\Lambda))^2. \quad (1)$$

Step 5. Use conditional variance results obtained in Step 4, equation (1), to obtain an unconditional variance estimate for y in the following manner:

$$E(y) = E_\Lambda(E(y|\Lambda))$$

$$\mathrm{var}(y) = E(\mathrm{var}(y|\Lambda)) + \mathrm{var}(E(y|\Lambda)). \quad (2)$$

Equation (2) is the variance of the distribution for the simulated data that incorporates uncertainty in the model parameters. In cancer microsimulation, we are often dealing with an output of counts (for example, cancer incidence, mortality, or number detected by screening); in this case the assumption that the outcome is a Poisson random variable is often reasonable and we can use a Poisson regression with overdispersion[29] to fit a hypersurface to the simulation results.

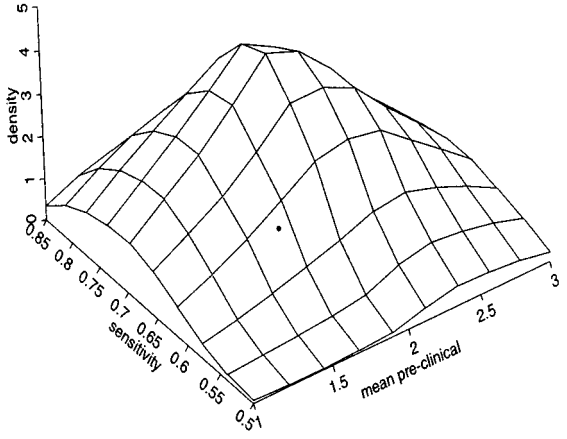

Figure 1. Multivariate distribution for Λ. Point represents maximum likelihood point estimate for the parameters

3. COMPARISON OF METHODS

We use a sample simulation with 100 simulation runs to compare an analysis that assumes no uncertainty, a parametric bootstrap approach and the proposed method described above. We then perform a bootstrap analysis with 1000 simulation runs which we take to be the gold standard for our example. Our intention is that this sample simulation demonstrates the procedures, rather than truly investigates the introduction of a cancer screening programme.

The simulation considers the introduction of a hypothetical breast cancer screening programme. A cohort of women, whose age at the start of the study is uniformly distributed between 40 and 50 years, are followed for a period of 30 years. Deaths from other causes are based on life table data with 10 year intervals. Incidence rates of pre-clinical disease are based on Surveillance, Epidemiology, and End Result (SEER) data for breast cancer incidence shifted back 5 years. We assume that the length of the pre-clinical period (defined as the time from the initial onset of disease to the time clinical symptoms appear) is exponential. We further assume that the disease is detectable by screening during this pre-clinical period. The screening test is performed every 2 years over the 30 year follow-up period. For simplicity, the main outcome of interest is the number of cases detected by screening. In this example, the parameter estimates that contain considerable uncertainty are the mean length of the pre-clinical period and the sensitivity of the screening test.

Let Λ represent a 2×1 vector containing estimates for the model parameters (average length of pre-clinical period, in years, and test sensitivity). Assume that Λ follows the multivariate normal distribution shown in equation (3). The mean lead time is strictly greater than zero and the sensitivity must be between zero and one. Therefore, we use a truncated multivariate normal distribution which includes only feasible values:

$$\Lambda \sim \mathrm{MVN}\left(\begin{pmatrix} 2{\cdot}0 \\ 0{\cdot}70 \end{pmatrix}, \begin{pmatrix} 0{\cdot}25 & -0{\cdot}0175 \\ -0{\cdot}0175 & 0{\cdot}0025 \end{pmatrix}\right). \tag{3}$$

This distribution is based on the reported results of a MISCAN simulation for breast cancer screening.[5] Figure 1 shows the distribution for Λ. The single point estimate (the mean of Λ)

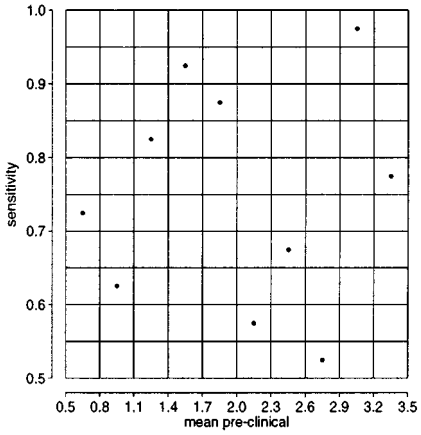

Figure 2. Latin square sample

represents the parameter values that we would use if we were to ignore the uncertainty in the model parameters and perform all runs using this single set of values.

3.1. No parameter uncertainty

Ignoring the uncertainty in the model parameters, we performed 100 simulations consisting of 10,000 life histories each using the parameter means ($\Lambda = (^{2}_{0{\cdot}70})$). This gave an overall simulated mean of 162 cases detected by screening and variance of 194.

3.2. Parametric bootstrap analysis

We performed a parametric bootstrap analysis for the simulation described by generating 100 points (that is, 100 choices for the mean pre-clinical length and sensitivity) from the distribution for Λ and running a simulation of 10,000 life histories a single time at each point. Empirical estimates for the mean and variance of the simulated number of cases detected by the screening were 162 and 335, respectively. We use the same number of runs as in the other methods for purposes of comparison. In general, 100 sampled points would be considered adequate to estimate standard deviation. However, due to the added complexity of running a stochastic simulation with each bootstrap parameter estimate, a sample size of 100 points may not be adequate in this situation. As the number of runs increases, the results from the bootstrap analysis approaches the true variance.

3.3. Parameter sampling design

We start by dividing the parameter space for Λ into a grid of squares of equal area and sample squares using a Latin squares design. We sample the first square uniformly from all columns and rows of the grid, we delete the associated column and row of the selected square from the grid, and then select the next square uniformly from the remaining columns and rows. We used parameter values at the middle of the squares to run 10 simulations for each sampled grid square, giving a total of 100 simulation runs. Figure 2 shows the points selected in our analysis.

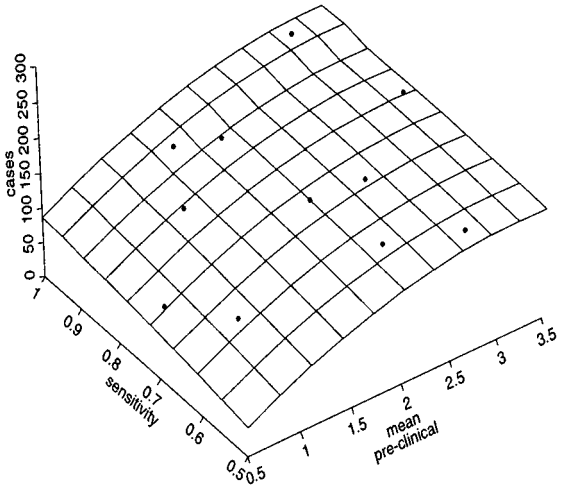

Figure 3. Fitted hypersurface

We fit a response hypersurface to the 100 simulated results using a Poisson regression (shown in Figure 3). The regerssion equation used appears in equation (4), where λ_1 is the mean lead time and λ_2 is the sensitivity of the screening test. Inclusion of a product term ($\lambda_1\lambda_2$) and a squared term for test sensitivity did not significantly improve the fit of the model.

$$y|\lambda_1, \lambda_2 \sim \text{Poisson}(\mu(\lambda_1, \lambda_2))$$
$$\ln(\mu(\lambda_1, \lambda_2)) = 3{\cdot}257 + 0{\cdot}861\lambda_1 + 0{\cdot}922\lambda_2 - 0{\cdot}137\lambda_1^2. \quad (4)$$

We tested for overdispersion using the methods described in Dean and Lawless[30] and found little evidence of overdispersion. We then integrate the response hypersurface with respect to the probability measure for Λ, giving $E_\Lambda(E(y|\Lambda)) = 157$ and $\text{var}_\Lambda(E(y|\Lambda)) = 395$ as described in Step 4 of our procedure. Step 5 gives an unconditional estimate of 552 for the variance of the number of cases detected.

3.4. Results

Table I shows results from the three different methods discussed. We then ran the parametric bootstrap analysis for an additional 900 runs, for total of 1000 runs. We consider this the gold standard for comparing the three methods discussed. Table II shows the results for the bootstrap of 1000 points. The confidence interval was based on percentile estimates from the simulated outcome. The method assuming no uncertainty and the bootstrap method with 100 simulation runs underestimated the standard deviation compared to the estimates based on 1000 simulation runs. The bootstrap method based on only 100 points could be sensitive to the parameters values generated, while the sampling design approach covers the parameter space. The sampling design approach relies on the regression equation representing the relationship between model parameter values and simulation output.

Table I. Estimates based on 100 simulation runs

	$E(Y)$	$V(Y)$	$SD(Y)$
1. No uncertainty	162	194	13·9
2. Bootstrap	162	335	18·3
3. Our method	157	552	23·5

Table II. Estimates based on 1000 simulation runs

	$E(Y)$	$V(Y)$	$SD(Y)$	95% CI
Bootstrap	158·3	497	22·30	(107, 198)

4. APPLICATION: PSA SCREENING AND PROSTATE CANCER MORTALITY RATES

In related work, we have developed a stochastic simulation model of PSA screening for prostate cancer with the goal of identifying natural history and screening efficacy parameters consistent with disease-specific mortality trends in the population. This reflects the model fitting goal of the simulation approach. In this case, the model consists of two main components, the first being identification of screen-detected cases and the second being the modelling of disease-specific mortality of these cases with and without screening. The first component depends on screening utilization and cancer rates, for which there are available data. The second depends on assumptions about the time of advancement of diagnosis due to screening (lead time) and the survival benefits of screen-detection. Different assumptions about these quantities lead to different estimates of the impact of screening on prostate cancer mortality trends.

In recent years, mortality rates for prostate cancer have levelled off and have begun to decline.[31] We can use the microsimulation developed to investigate the possibility that recent observed trends are a result of prostate cancer screening. One approach is to use the simulation model to predict the effect of screening based on parameter values considered likely. A second approach is to identify a set of parameter values that are consistent with the hypothesis that recent declines are due to screening. Our method facilitates both of these approaches by the fitting of a hypersurface and incorporating the prior probability distribution for the model parameters.

The simulation begins by generating a cohort of men aged 65 to 84 in 1988. We refer to this group as the ageing cohort, since their age advances one year with each year of the simulation. The age distribution in 1988 is determined by U.S. census data; we assume exact age is uniformly distributed within five-year age intervals. For each individual, we generate an age at death due to causes other than prostate cancer from U.S. life tables.[32] Men exit the simulation at the age of 85 or at the time of death due to other causes, whichever comes first.

Each year, from 1988 to 1994, we generate an indicator of screening utilization based on rates of PSA screening observed among Medicare enrollees (Table III).[33] Individuals still alive are eligible for a first screen if they have not had a previous screen, and they are eligible for a second or later screen if they have had at least one previous screen and have not been diagnosed with cancer. We assume that scheduled time of screening is uniformly distributed within one year. We calculate disease status among screened individuals based on cancer detection rates from

Table III. PSA screening rates

	Year						
	1988	1989	1990	1991	1992	1993	1994
Ages	65–84	66–84	67–84	68–84	69–84	70–84	71–84
Percent 1st screen	1·297	2·929	4·247	12·037	13·802	10·430	6·724
Percent at least 1 screen	1·297	3·498	6·041	15·958	24·007	27·136	27·517

Medicare data regarding men who received PSA tests from January 1988 to December 1993.[33] The likelihood of cancer detection following a PSA test is estimated as 3 per cent for individuals screened for the first time and 2 per cent for later screens.[33]

The screen-detected individuals are the focus of the model since they are the only cases whose prostate cancer mortality might be affected by screening. For each screen-detected case, the program computes the data of death from prostate cancer had screening not taken place and the date of death from prostate cancer given detection by screening. These computations depend on the average lead time and the survival benefit due to screening as follows. Let T_1 be the year of death without screening, and T_2 the year of death given screen-detection. Then $T_1 = t_d + l_t + s_1$ and $T_2 = t_d + l_t + s_2$, where t_d is the time of screen detection, l_t is the lead time, and s_1 and s_2 are the times to prostate cancer death without and with screening, respectively. We assume that the lead time follows an exponential distribution. The time to prostate cancer death without screening is based on SEER relative survival of prostate cancer cases diagnosed from 1980 to 1987. These years were selected to represent a relatively recent calendar time period prior to the introduction of screening in the population. The relative survival depends on age within five-year age intervals.

We assume that screening affects survival through a proportional hazards model[34] (that is, each year the prostate cancer mortality rate with screening is a factor r (less than 1) times the prostate cancer mortality rate without screening). Both r and the mean lead time are input by the user. Note that we model survival in either case from the original time of diagnosis without screening, therefore we assume that the screen-detected cases live at least until this point.

We computed the number of deaths with and without screening for 1994; their difference gives the number of deaths prevented by screening. We also represent deaths prevented as a percent of those expected without screening. We present results for the age group 71–84, since this was the largest group present in all years.

Note that all results are based on the time of prostate cancer death with and without screening rather than survival from diagnosis. Consequently, lead time bias is avoided. However, length bias is not explicitly considered. In fact, the simulation explicitly focuses on screen-detected cases, which we may expect show better survival than SEER cases diagnosed before 1987, even in the absence of screening. Since no quantitative information is available on how much better their prognosis might be, we have not modelled this phenomenon explicitly. Therefore, it is possible that the model's prostate cancer mortality without screening may overestimate the expected mortality.

There is considerable uncertainty related to the mean lead time associated with the screening test and the hazard ratio for survival of individuals detected by screening versus those detected through symptoms. We assume a multivariate normal distribution truncated to include only

non-negative values for the mean lead time and hazard ratio. The mean lead time parameter of 7 years is based on the work of Whittemore *et al.*[35] and Morrell *et al.*[36] Whittemore *et al.* showed that PSA level was a sensitive marker for men who would develop prostate cancer within the next 7 years based on longitudinal data obtained retrospectively. Morrell *et al.* fit a changepoint model to longitudinal PSA readings of men who developed prostate cancer and found that PSA levels started increasing on average 6·4 years for those diagnosed with local or regional cancer and 10·72 years for those diagnosed with metastatic cancer. Unfortunately, there is little information available on the hazard ratio. Therefore, the distribution of hazard ratio is based on screening tests from other cancer sites (for example, mammography) and our prior belief of what is a reasonable range of values:

$$\Lambda \sim \mathrm{MVN}\left(\begin{pmatrix}7{\cdot}0\\0{\cdot}6\end{pmatrix},\begin{pmatrix}1 & -0{\cdot}035\\-0{\cdot}035 & 0{\cdot}0025\end{pmatrix}\right).$$

As in the previous example, we performed 10 microsimulation runs at each set of parameter values (100 total simulation runs) and a Poisson regression was used to fit a hypersurface to the simulation output of number of deaths prevented. Equation (5) shows that regression equation, where λ_1 is the mean lead time and λ_2 is the hazard ratio:

$$\begin{aligned} y|\,\lambda_1,\lambda_2 &\sim \mathrm{Poisson}(\mu(\lambda_1,\lambda_2))\\ \ln(\mu(\lambda_1,\lambda_2)) &= 17{\cdot}013 - 3{\cdot}864\lambda_1 - 7{\cdot}905\lambda_2 + 0{\cdot}451\lambda_1^2 - 0{\cdot}020\lambda_1^3 + 0{\cdot}827\lambda_1\lambda_2. \end{aligned} \tag{5}$$

Unlike the previous example, this application included a large amount of overdispersion. The Poisson regression estimated the overdispersion parameter as 10·01, so that the $\mathrm{var}(y) = 10{\cdot}01E(y)$.

Figure 4 shows the large amount of variability at each set of parameter values. The graphs plot the number of deaths prevented generated by each of the 10 simulation runs for each set of parameters. The solid lines show the estimated mean based on the Poisson regression. The Poisson regression predicts the mean of the 10 values rather well, but does a poor job of predicting the results of a single simulation run due to the extra Poisson variation as reflected by the large overdispersion parameter.

The points sampled from the Latin square are shown in the left of Figure 5, and the fitted hypersurface is shown at the right. The mean of the 10 simulation runs at each set of parameter values are also plotted with the hypersurface.

We then integrate the response surface with respect to the multivariate probability distribution of the model parameters to get estimates of the mean and variance of the response surface as described in Step 4 of the suggested procedure:

$$E_\Lambda(E(y|\Lambda)) = \int_\Lambda g(\Lambda) f(\Lambda) \mathrm{d}\Lambda = 49$$

$$\mathrm{var}_\Lambda(E(y|\Lambda)) = \int_\Lambda g(\Lambda)^2 f(\Lambda) \mathrm{d}\Lambda - E_\Lambda(E(y|\Lambda))^2 = 23.$$

Finally we calculated the unconditional variance for the number of deaths prevented through screening:

$$\mathrm{var}(y) = E(\mathrm{var}(y|\Lambda)) + \mathrm{var}(E(y|\Lambda)) = 10{\cdot}01 \times 49 + 23 = 519$$

$$\text{Standard deviation}(y) = 22{\cdot}8.$$

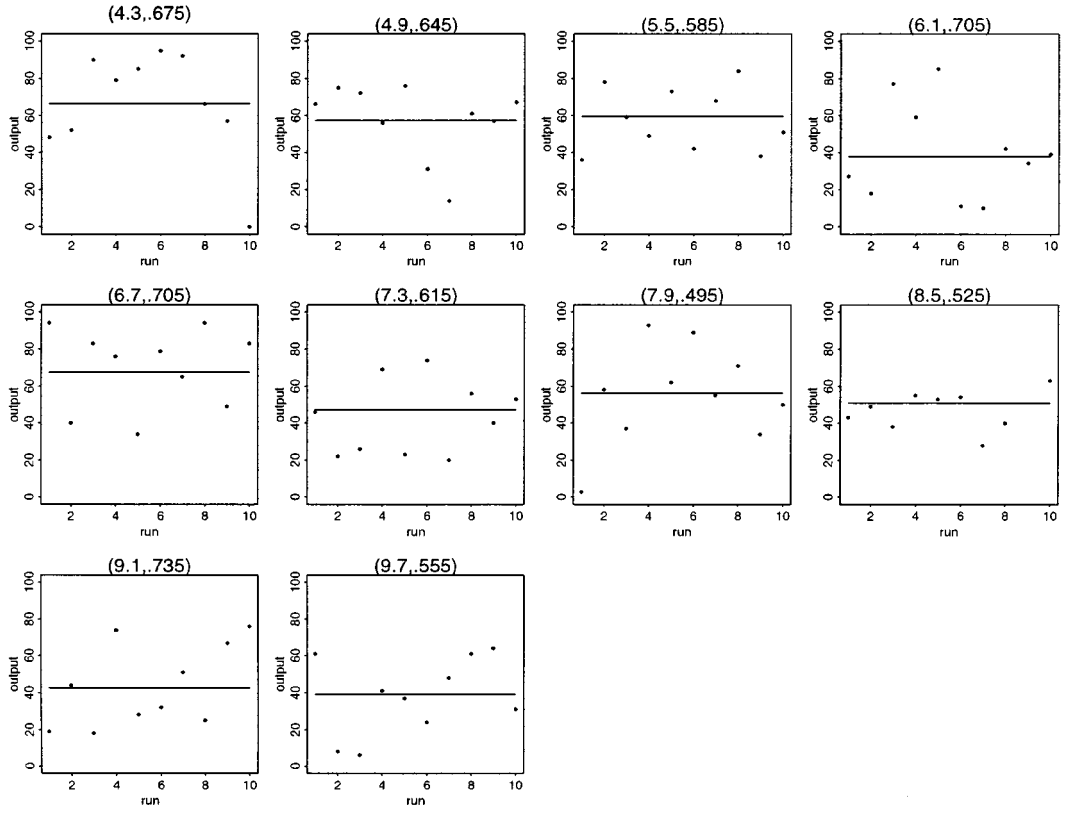

Figure 4. Simulation results for the parameter values sampled, line represents the mean of the 10 simulation runs at a given set of parameter values

In this application, stochastic variation is the major influence on the output variance of the simulated screening benefit, relatively speaking. Proportionally, the additional variation due to uncertainty in the model parameters contributes only a small part to the overall variation.

The number of deaths prevented corresponds to approximately a 0·03 percent reduction in mortality conditioned on our prior probability distribution for the model parameters. The results suggest that recent prostate cancer mortality trends are probably not attributable to PSA screening. Further, the response surface in Figure 5 shows that screening could have accounted for observed change in mortality only at very low values for mean lead time and low values for hazard ratio, believed to be unlikely.

5. DISCUSSION

The use of microsimulation has allowed researchers to study complex processes and relationships that are not possible with classical statistical methods. The growing literature on microsimulations and their increased popularity reflects advances in statistical theory as well as computational methods and resources. This field represents an exciting intersection of statistics, computer

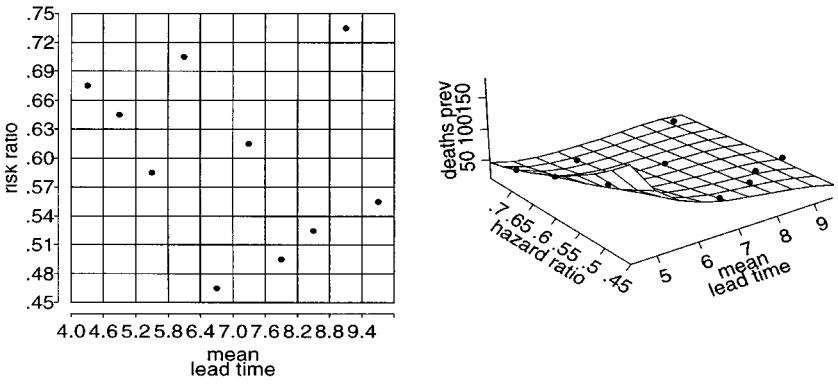

Figure 5. Latin square sample and fitted hypersurface for application

science, economic decision theory and clinical sciences. However, a danger of the use of such complex, sophisticated methods is the failure to appreciate fully the degree of uncertainty associated with the results. By now, the need for sensitivity analyses is well recognized by most researchers who use microsimulations. The method we propose allows researchers to maintain this perspective. By fitting and graphing a response surface, we preserve the essence of the sensitivity analysis. As a number of authors have pointed out, sensitivity analysis falls short on several counts.[18,22,37] A sensitivity analysis provides equal weighting of the extremes of the parameter space. This can lead to an exaggeration of the effects of uncertainty when there is little reason to believe that the parameters will attain such extreme values. Another shortcoming of sensitivity analysis is that it does not produce a single number that summarizes the uncertainty associated with an estimate. This makes direct comparison of model outcomes and observed data difficult. Finally, many sensitivity analyses are done in a univariate manner, first varying values of one parameter and then another. Although some authors recommend two- or multi-way sensitivity analyses, this can be challenging in practice.

The method we propose addresses each of these issues. It uses a multivariate probability distribution to take into account correlation between model parameters. By combining the probability distribution of the model parameters and the response surface, it produces a single estimate of the mean and variance for the quantity under study that incorporates the effects of parameter uncertainty. Finally, this variance may not be as large as one might expect from examining the sensitivity analyses alone since improbable extremes are down-weighted.

The two examples we consider provide an interesting contrast and highlight some strengths of the approach we propose. In the first simulation of a breast cancer intervention, overdispersion was not significant and the magnitude of the variance conditional on the parameter values was small relative to the variation in the estimated mean values on the hypersurface. The reverse was true for the second example involving PSA and prostate cancer mortality. This is critical information for those who perform and interpret the results of these microsimulations.

Besides providing useful information concerning the variation in the process, another strength of this approach is the decoupling of the hypersurface estimation and the probability distribution of the hyperparameters. Unlike the parametric bootstrap method which concentrates model

parameter value selections in specified regions of high probability, our method allows researchers to select their own probability distribution of the model parameters, if they so choose, and to compute their own estimates of the mean and variance.

In addition to reviewing and proposing methods for incorporating uncertainty when reporting microsimulation results, this paper has also delineated the objectives of microsimulation in the context of cancer research. It has elucidated the role of microsimulation in simulating randomized clinical trials, and foresees a larger role of microsimulation in surveillance modelling of population dynamics. The latter objective suggests that we broaden our notion of how one presents and analyses microsimulation results. In surveillance modelling, the emphasis is less on the absolute level or magnitude of an effect and more on interpretation of trend patterns. Methods being developed for surveillance modelling, such as overdispersed Poisson regression with join points, might reasonably apply to microsimulation. In this case, inferences may centre around trends in cancer incidence rates that include join point locations.

The approach developed in this article represents a significant advance over what is currently available to those researchers who face the task of coherently and honestly presenting results from a simulation of a real process with considerable parameter uncertainty. We plan to expand our work to include the uncertainty of the model structure as another level in our procedure.

ACKNOWLEDGEMENTS

Research supported, in part, by contract NCI N01-CN-05230. The authors would like to thank Michael McKay from Los Alamos National Laboratory and Craig Volonoski from Onward Inc. for reviewing our work, providing helpful suggestions, and pointing our similar work in other areas of research.

REFERENCES

1. Zelen, M. and Feinleib, M. 'On the theory of screening for chronic disease', *Biometrika*, **56**, 601–614 (1969).
2. Zelen, M. 'Optimal scheduling of examinations for the early detection of disease', *Biometrika*, **80**, 279–293 (1993).
3. Duffy, S. W., Chen, H. H., Tabar, L. and Day, N. E. 'Estimation of mean sojourn time in breast cancer screening using Markov chain model of both entry to and exit from the preclinical detectable phase', *Statistics in Medicine*, **14**, 1531–1543 (1995).
4. Habbema, J. D. F., van Oortmarssen, G. J., Lubbe, J. T. N. and van der Maas, P. J. 'The MISCAN simulation program for the evaluation of screening for disease', *Computer Methods and Programs in Biomedicine*, **20**, 79–83 (1984).
5. van Oortmarssen, G. *Evaluation of Mass Screening for Cancer a Model-based Approach*, Offsetdrukkerij, Alblasserdam, The Netherlands, 1995.
6. Loeve, F., van Ballegooijen, M., Boer, R. and Habbema, J. D. F. 'Report for working meeting Miscan colorectal cancer screening model', Erasmus University Rotterdam, Project number N01-CN-55186, 1997.
7. Etzioni, R. and Thompson, I. M. 'Prostate cancer and computer models: background, limitations and potential', *Urological Oncology*, **2**, 57–64 (1996).
8. Urban, N., Drescher, C., Etzioni, R. and Colby, C. 'Use of a stochastic simulation model to identify an efficient protocol for ovarian cancer screening', *Controlled Clinical Trials*, in press.
9. Szeto, K. L. and Devlin, N. J. 'The cost-effectiveness of mammography screening: evidence from a microsimulation model for New Zealand', *Health Policy*, **38**, 101–115 (1996).
10. McKay, M. D. *Evaluating Prediction Uncertainty*, Report NUREG/CR-6311, LA-12915-MS, U.S. Nuclear Regulatory Commission, 1995.

11. McKay, M. D., Wendelberger, J. R. and Morrison, J. D. 'Model uncertainty in stochastic simulation models; contributions to the development of a statistical framework for joint mobility modeling', Working paper LA-UR-97-1366, Los Alamos National Laboratory, 1997.
12. Ackerman, E. 'Simulation of micropopulations in epidemiology: Tutorial 3. Simulation model evaluation methods', *International Journal of Biomedical Computing*, **37**, 195–204 (1994).
13. Ackerman, E. 'Simulation of micropopulations in epidemiology: Tutorial 1. Simulation: an introduction', *International Journal of Biomedical Computing*, **36**, 229–238 (1994).
14. Ackerman, E. 'Simulation of micropopulations in epidemiology: Tutorial 2. Analytic forms of event probabilities', *International Journal of Biomedical Computing*, **37**, 139–149 (1994).
15. Ackerman, E. 'Simulation of micropopulations in epidemiology: Tutorial 4. Evaluations of simulation models', *International Journal of Biomedical Computing*, **39**, 219–229 (1995).
16. Ma, J. Z., Ackerman, E. and Yang, J. 'Parameter sensitivity of a model of viral epidemics simulated with Monte Carlo techniques. I. Illness attack rates', *International Journal of Biomedical Computing*, **32**, 237–253 (1993).
17. Ma, J. Z. and Ackerman, E. 'Parameter sensitivity of a model of viral epidemics simulated with Monte Carlo techniques. II. Durations and peaks', *International Journal of Biomedical Computing*, **32**, 255–268 (1993).
18. Doubilet, P., Begg, C. B., Weinstein, M. C., Braun, P. and McNeil, B. J. 'Probabilistic sensitivity analysis using Monte Carlo simulation a practical approach', *Medical Decision Making*, **5**, 157–177 (1985).
19. Faraway, J. J. 'On the cost of data analysis', *Journal of Computational and Graphical Statistics*, **1**, 213–229 (1992).
20. Draper, D. 'Assessment and propagation of model uncertainty', *Journal of the Royal Statistical Society, Series B*, **57**, 45–97 (1995).
21. Madigan, D., Raftery, A. E., Volinsky, C. T. and Hoeting, J. A. 'Bayesian model averaging', AAAI, Workshop on Integrating Multiple Learned Models, 1996.
22. Manning, W. G., Fryback, D. G. and Weinstein, M. C. 'Reflecting uncertainty in cost-effectiveness analysis', *in* Gold, M. R., Siegel, J. E., Russell, L. B. and Weinstein, M. C. (eds), *Cost-effectiveness in Health and Medicine*, Oxford University Press, New York, 1996.
23. Hodges, J. S. 'Uncertainty, policy analysis and statistics', *Statistical Science*, **2**, 259–291 (1987).
24. Downing, D. J., Gardner, R. H. and Hoffman, F. O. 'An examination of response-surface methodologies for uncertainty analysis in assessment models', *Technometrics*, **27**, 151–163 (1985).
25. Helton, J. C., Johnson, J. D., McKay, M. D., Shiver, A. W. and Sprung, J. L. 'Robustness of an uncertainty and sensitivity analysis of early exposure results with the MACCS reactor accident consequence model', *Reliability Engineering and System Safety*, **48**, 129–148 (1995).
26. Cole, P. and Rodu, B. 'Declining cancer mortality in the United States [see comments]', *Cancer*, **78**, 2045–2048 (1996).
27. Efron, B. *The Jackknife, the Bootstrap and other Resampling Plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, **38**, 1982.
28. McKay, M. D., Beckman, R. J. and Conover, W. J. 'A comparison of three methods for selecting values of input variables in the analysis of output from a computer code', *Technometrics*, **21**, 239–245 (1979).
29. Hakulinen, T. and Dyba, T. 'Precision of incidence predictions based on Poisson distributed observations', *Statistics in Medicine*, **13**, 1513–1523 (1994).
30. Dean, C. and Lawless, F. 'Tests for detecting overdispersion in Poisson regression models', *Journal of the American Statistical Association*, **84**, 467–472 (1989).
31. Ries, L. A. G., Kosary, C. L., Hankey, B. F., Miller, B. A., Harras, A. and Edwards, B. K. (eds) *SEER Cancer Statistics Review, 1973–1994*, National Cancer Institute, NIH Pub. No. 97-2789, Bethesda, MD, 1997.
32. National Center for Health Statistics. *Vital Statistics of the United States, 1986, Vol. II Mortality Part A*, DHHS Pub. No. (PHS) 88-1122, Public Health Service, Washington, US Government Printing Office, 1988.
33. Legler, J., Feuer, E., Potosky, A., Merrill, R. and Kramer, B. 'The role of prostate-specific antigen (PSA) in the recent prostate cancer incidence decline in the USA', submitted for publication.
34. Miller, R. G. *Survival Analysis*, Wiley, New York, 1981.

35. Whittemore, A. S., Lele, C., Friedman, G. D., Stamey, T., Vogelman, J. H. and Orentreich, N. 'Prostate-specific antigen as predictor of prostate cancer in black men and white men', *Journal of the National Cancer Institute*, **87**, 354–360 (1995).
36. Morrell, C. H., Pearson, J. D., Carter, H. B. and Brant, L. J. 'Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer', *Journal of the American Statistical Association* **90**, 45–53 (1995).
37. O'Brien, B. J., Drummond, M. F., Labelle, R. J. and Willan, A. 'In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care', *Medical Care*, **32**, 150–163 (1994).

