Effects of Exposure Uncertainties in the TSCE Model and Application to the Colorado Miners Data

Author(s): Wolfgang F. Heidenreich, E. Georg Luebeck, Suresh H. Moolgavkar
Source: Radiation Research, 161(1):72-81.
Published By: Radiation Research Society
https://doi.org/10.1667/RR3089
URL: http://www.bioone.org/doi/full/10.1667/RR3089

RADIATION RESEARCH **161**, 72–81 (2004)
0033-7587/04 $15.00

Effects of Exposure Uncertainties in the TSCE Model and Application to the Colorado Miners Data

Wolfgang F. Heidenreich,[a,1] E. Georg Luebeck[b] and Suresh H. Moolgavkar[b]

[a] *GSF—National Research Center for Environment and Health Institute for Radiation Protection, 85764 Neuherberg, Germany; and*
[b] *Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, Washington 98109-1024*

Heidenreich, W. F., Luebeck, E. G. and Moolgavkar, S. H. Effects of Exposure Uncertainties in the TSCE Model and Application to the Colorado Miners Data. *Radiat. Res.* 161, 72–81 (2004).

The simulations in this paper show that exposure measurement error affects the parameter estimates of the biologically motivated two-stage clonal expansion (TSCE) model. For both Berkson and classical error models, we show that likelihood-based techniques of correction work reliably. For classical errors, the distribution of true exposures needs to be known or estimated in addition to the distribution of recorded exposures conditional on true exposures. Usually the exposure uncertainty biases the model parameters toward the null and underestimates the precision. But when several parameters are allowed to be dependent on exposure, e.g. initiation and promotion, then their relative importance is also influenced, and more complicated effects of exposure uncertainty can occur. The application part of this paper shows for two different types of Berkson errors that a recent analysis of the data for the Colorado plateau miners with the TSCE model is not changed substantially when correcting for such errors. Specifically, the conjectured promoting action of radon remains as the dominant radiation effect for explaining these data. The estimated promoting action of radon increases by a factor of up to 1.2 for the largest assumed exposure uncertainties.

INTRODUCTION

Traditional statistical analyses of epidemiological data and of data from animal experiments use the best available information on exposure. It is known, however, that exposures may be measured with error, either systematic or random or both. Correlated errors may also occur when the same uncertain quantity is used in the calculation of several exposures. Exposure measurement error may affect not only the estimates of parameters but their confidence intervals as well (*1*). In most simple applications, ignoring exposure uncertainties leads to an underestimation of risk. Scientists working in radiation epidemiology are aware of the problem (*2*, *3*), and much work has been done to take account of dose uncertainty for quantitative risk estimates in occupational cohorts (*5*, *7*), in the A-bomb survivors (*6*), for residential radon (*8*), and by simulation (*4*). The proceedings of a recent workshop (*9*) give an up-to-date overview. In the statistical literature, several methods for taking into account various forms of error structures have been proposed. A useful summary can be found in ref. (*1*).

The two-stage clonal expansion (TSCE) model is a generalization of Knudson's recessive oncogenesis model (*10*); it can be thought of as a mathematical formalization of the initiation-promotion-progression paradigm of carcinogenesis. Exact solutions for the hazard function and the survival function of the model have been found for piecewise constant exposures (*11*), which can be solved by fast recursive calculations (*12*). The model describes well a wide range of epidemiological data (*13–15*) and data from animal experiments (*16–18*). In some of these applications, it was concluded that the TSCE model can supplement epidemiological analysis in risk estimates.

Especially when extrapolations are needed to exposure schemes for which no direct data are available, e.g. for low doses and low dose rates, the mechanistic foundation of the TSCE model can add to the credibility of the predictions. In all applications of the TSCE model to data, the assessed exposures were assumed to be accurate. Due to the fast recursive computer code mentioned above, it is now feasible to account explicitly for exposure uncertainty.

Here we examine the impact of exposure uncertainty on the estimated parameters of the TSCE model. We consider both classical and Berkson error structures but restrict our discussion to random errors without correlations.

In the classical error model, the recorded exposures, w, fluctuate randomly around the true exposure, x (see Fig. 1). This applies, for example, when the true exposure is measured inaccurately. The fluctuation can be additive, $W = X + U$; in this case, W is an unbiased measure of X if U has expectation 0. The error structure may be multiplicative, $W = X \times U$; then W is an unbiased estimate of X if the expectation of U is 1.

In the Berkson error model, the true exposures x fluctuate

[1] Address for correspondence: GSF—National Research Center for Environment and Health Institute for Radiation Protection, 85764 Neuherberg, Germany; e-mail: heidenreich@gsf.de.

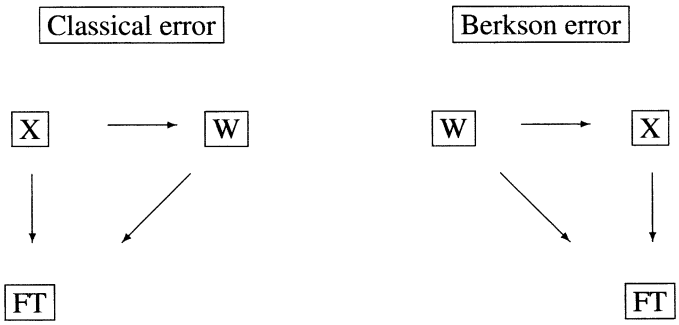

FIG. 1. Diagrams of classical and Berkson errors. X denotes the true exposure, W the recorded exposure. Horizontal arrows indicate the error model, vertical arrows the generation of failure times, FT, and slanted arrows the raw analysis.

randomly around the recorded exposures *w*. This applies, for example, when an experimenter aims at a predefined value of exposure, or when precise measurements at central monitors are made, but the true exposures to individuals fluctuate due to differences in behavior or in physiology. Again an additive model with $X = W + U$ or a multiplicative model with $X = W \times U$ is possible.

We show by simulation that our likelihood-based methods work well. As a concrete application, we apply these methods to analyses of the Colorado plateau miner cohort (*15*). In that part of the work, the main question to be answered is whether the suggested strong promoting action of radon might be in part an artifact of errors in dosimetry.

MATERIALS AND METHODS

The TSCE Model

A sketch of the TSCE model is given in Fig. 2. Normal healthy cells (whose number N usually is not known) mutate with rate μ_1 to intermediate cells. Thus these are created with the initiation rate $\nu = N\mu_1$. They can divide into two intermediate cells with rate α, die or differentiate with rate β, or divide into an intermediate cell and a malignant cell with the transformation rate μ. The progression from a malignant cell to an observable tumor is usually described with a lag time. Not all of these parameters can be determined from data. Therefore, identifiable parameters are used here. The notation and the shape of the exposure response are similar to those in ref. (*15*).

An explicit form for the spontaneous hazard is (*19*)

$$h^s(t) = \frac{y_0[e^{(g_0+2q_0)t} - 1]}{g_0 + q_0[e^{(g_0+2q_0)t} + 1]}. \tag{1}$$

It depends on three parameters that are chosen to determine characteristic features of the hazard function. The values of the parameters are close to those found for the background parameters in analyses of the atomic bomb survivors (*13*), adjusted such that they give about 10% tumor cases in the simulated data set. Explicitly for the product of spontaneous initiation and transformation rates, $y_0 = 10^{-6}$ (year^{-2}) is used, for the effective clonal expansion rate $g_0 \equiv \alpha_0 - \beta_0 - \mu_0 = 0.11$ (year^{-1}), and for the stochasticity parameter, $q_0 = 10^{-4}$ (year^{-1}). This latter parameter is a more complicated function of the biological parameters chosen such that for asymptotically high age the hazard takes the constant values y_0/q_0.

The action of exposure rate d is assumed on the initiation rate, the transformation rate, the effective clonal expansion rate (promotion), or combinations of these. The dependence on exposure rate is assumed to be linear for the mutational events initiation [$\nu(d) = \nu_0(1 + y_1 d)$] and transformation [$\mu(d) = \mu_0(1 + m_1 d$] and to be initially linear up to a constant level for promotion. This gives, for the chosen identifiable parameters,

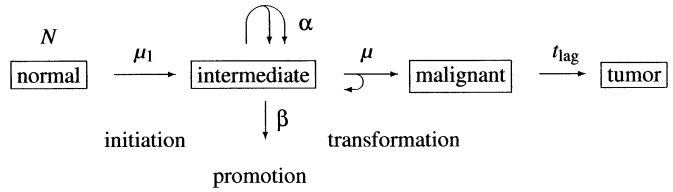

FIG. 2. Sketch of the TSCE model.

$$\begin{aligned}
&\text{Initiation:} && y(d) \equiv \nu(d)\mu_0 = y_0(1 + \boldsymbol{y}_1 d),\\
&\text{Promotion:} && g(d) \equiv \alpha(d) - \beta(d) - \mu(d)\\
&&& \quad = g_0 + \min(\boldsymbol{g}_{\text{lin}} d, \boldsymbol{g}_{\text{level}}),\\
&\text{Transformation:} && m(d) \equiv \mu(d)/\mu_0 = 1 + \mathbf{m}_1 d.
\end{aligned} \tag{2}$$

The parameters in boldface are exposure-related parameters in the models. We use a lag time of 0. The values for these parameters are chosen such that about 10% of the simulated records develop an exposure-induced tumor, about the same proportion that occurs spontaneously. The selected values are given in the respective figures. The recursive formulas in ref. (*12*) are used for calculating the hazard and the survival functions.

Simulated Data

Data sets are usually simulated for 10,000 records. Twice that number was used to increase stability when both initiation and promotion were made dependent on dose. Only one exposure period is considered for simplicity. The exposure pattern is obtained by randomly selecting an age at start of exposure between 20 and 40 years, a duration between 0 and 20 years, and an exposure rate between 0 and 1, all distributed uniformly. When simulating classical error, this exposure rate is interpreted as true rate x; for Berkson error it is the recorded rate w.

Exposure uncertainties are included by multiplying the rate of exposure with samples from a log-normal distribution with mean 1 and shape parameter σ (see the Appendix for details). For an exposure rate of 1, the density is plotted in Fig. 3 for several levels of uncertainty. For classical

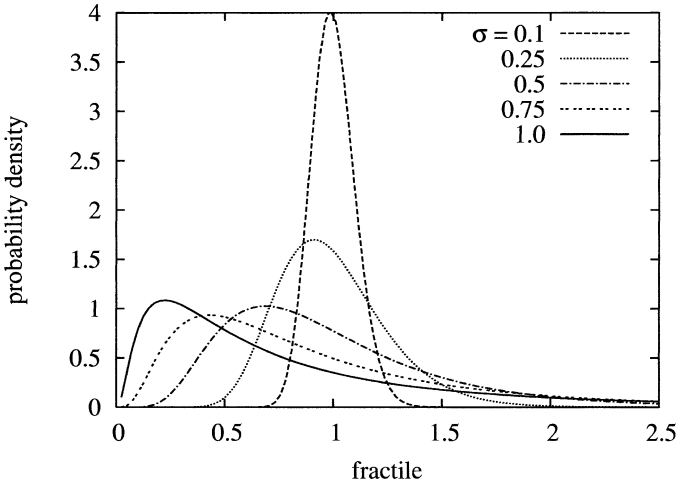

FIG. 3. Some probability densities for the exposure uncertainty as used in the text. They are log-normally distributed with the mean 1 and shape parameter σ.

errors, the recorded exposure w is calculated from the true exposure x, while for Berkson error the situation is reversed, as shown in Fig. 1.

Failure times are calculated using the TSCE model for the true exposures and an exposure-independent competing risk which describes the end of follow-up (called ''death'' here) for other reasons. Effectively two independent ages of death for each record are calculated randomly, one from the competing risk and one from the TSCE model. The earlier event gives the age at death and the cause of death. To minimize the computational burden, the following procedure is applied: The age of death t_c from the competing risk is obtained by randomly choosing t_c from a Gaussian distribution, centered at age 70, with a standard deviation of 15 years. Ages below 10 years are not accepted. With the TSCE model the survival probability for tumor $S(t_c)$ is calculated for this age. Also randomly, a survival probability S_T for the tumor is chosen from the evenly distributed numbers between 0 and 1. If $S(t_c) > S_T$, then the tumor appears later than the competing death, and t_c is the failure time. Otherwise the age at death due to tumor is calculated by bracketing, based on the value of S_T.

Likelihoods

The analyses of data with the TSCE model are usually done by maximizing the likelihood

$$L = \Pi l_w, \tag{3}$$

where the product is over all records and l_w is the likelihood contribution of an individual record. When no exposure uncertainty is taken into account, the recorded exposures w and the true exposures x are equal. Then the likelihood contribution is $l_w = l_x$, with

$$l_x = \begin{cases} h(t)S(t) & \text{if death is caused by tumor} \\ S(t) & \text{otherwise.} \end{cases} \tag{4}$$

Here $h(t)$ is the hazard function at age t of death, and $S(t)$ is the corresponding survival function. These are calculated using the recursion formulas in ref. (*12*). The negative log-likelihood is minimized using the variable-metric routine MIGRAD in the function minimizer program package MINUIT from CERN (*20*). The derivatives of the negative log-likelihood with respect to the estimated variables are calculated numerically. This is one of the standard ways to use MINUIT.

When the Berkson error situation applies, then the probability density $p(x|w)dx$ of the true exposure, given the recorded one, is known. The average likelihood contribution l_w of one record is (ref. *1*, p. 151)

$$l_w = \int l_x p(x|w)\, dx. \tag{5}$$

In this equation, l_x is the likelihood conditional on the particular value x of the true exposure history. When the distribution used to describe the error structure has more than one dimension, then multidimensional integrals are needed.

When the classical error situation applies, then the probability density $p(w|x)dw$ of the recorded exposure, given the true one, is known. The distribution of the exposures $p(x)dx$ is needed to calculated $p(x|w)dx$ (see also ref. *1*, p. 147). In the simulations this is no problem, since the distribution of true exposures is used to produce the data. Using the Bayes rule, the average likelihood contribution of one record can be expressed as

$$l_w = \frac{\int l_x p(w|x)p(x)\, dx}{c(w)}, \tag{6}$$

where $c(w) = \int p(w|x)p(x)dx$. The denominator $c(w)$ depends only on the recorded dose w and is not required for fitting.

For the log-normal distribution of errors which is used here, the integrals can be calculated efficiently using Gauss-Hermite integration, when only a one-dimensional uncertainty distribution is used. This situation applies in our simulations. For details, see the Appendix. The minimization of the negative log-likelihood is done in all cases exactly as described above.

Colorado Miners

The U.S. Public Health Service (USPHS) established a database on 3,347 white and 756 nonwhite male uranium miners working in the Colorado Plateau between 1950 and 1964. Recently the database was updated by NIOSH through at least the end of 1990. In addition to detailed information on the radon exposure, information on smoking was ascertained by a questionnaire in 1986. This data set was analyzed with the TSCE model in ref. (*15*), where more details can be found. The dependence on the radon exposure rate d_r of the initiation and promotion parameters in the model was very similar to what is used here in Eq. (2). Instead of the contribution $y_1 d$ to initiation used in the simulation here, $y_4 d_r$ was used in ref. (*15*), and the functional form of the exposure dependence of the clonal expansion rate (promotion) was

$$g(d_r) \equiv g_0[1 + g_2 \log(1 + g_1 d_r)]. \tag{7}$$

Additionally, a dependence on the smoking rate and a birth-year dependence of initiation were used.

The information on radon exposure in the data set is calculated from measurements in the mine shafts. Details of this procedure and possible levels of exposure uncertainties have been discussed recently (*7*). It is beyond the scope of this paper to use all of the information. Greatly simplified, the necessary steps are the following: The radon concentration was typically measured in the mines a few times a year; the errors of these measurements are of the classical type. The radon concentration has been extrapolated to other periods and between shafts; the uncertainties from these extrapolations are of the Berkson type. The individual exposures were calculated from the durations at the workplace and the estimated radon concentrations in the air of the mine shafts; uncertainties in these calculations are again of the Berkson type. Our interest in predominantly in relating true exposures to the centrally monitored measurements, and the Berkson error structure is therefore used. We acknowledge that a more precise model of exposure uncertainty could have a component of classical error.

The cumulative exposure of some miners based on the work history was compared in ref. (*21*) to lead-210 measurements in bones after the deaths of the miners. It was found to be ''within a standard error of $\pm$ 0.34 of the average of the two values''. We conclude that a shape parameter σ of the log-normal distribution between 0.5 and 1.0 is a plausible magnitude for the exposure uncertainties (see Fig. 3).

There are up to 17 intervals of exposure to smoking and radon in the data set. These time-varying patterns of exposure are modeled by piecewise constant exposures in the application of the TSCE model. Two greatly simplified types for exposure uncertainty are considered here: (1) All radon exposures of a miner are described by the same uncertainty factor, and (2) all intervals are treated independently with different factors. The truth is likely in between: A treatment of uncertainty depending on mine shaft and calendar year would be preferable. This would allow us to include the correlation which arises between the radon concentration error component of the Berkson error for all miners who worked in the same shaft at the same time. The standard error of this component is estimated at about 50% of the mean (*7*), which is well below our plausible magnitude for the range of total Berkson error.

For the first type of Berkson error, the integral in Eq. (5) can be performed using the Gauss-Hermite formalism. When all intervals are treated independently, a multi-integral has to be calculated. This is done by Monte Carlo integration with N (typically 1000) sample points x_j for each miner. When they are taken randomly from the density $p(x|w)dx$ by using Eq. (A1), then the integral (Eq. 5) is approximated by the mean value of the likelihood contribution l_{x_j} (ref. *22*, p. 306ff). Thus the likelihood contribution for each miner is calculated with

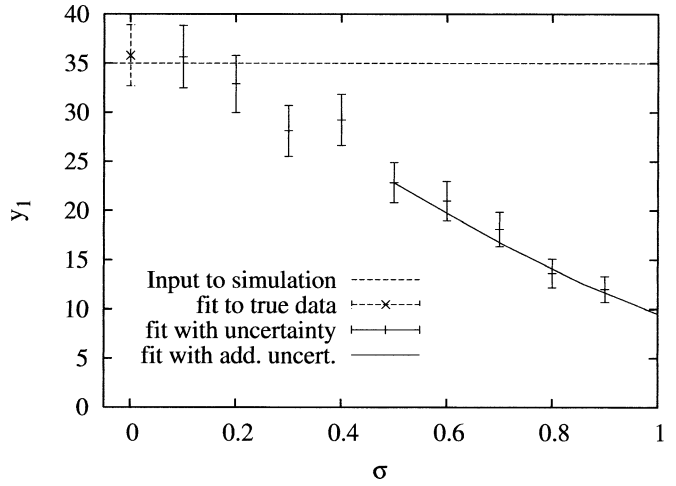

FIG. 4. Effect of classical exposure uncertainty on the initiation parameter y_1 (see Eq. 2). Shown are the results for a simulated data set with 10,000 records and failure times based on the true exposure. The data are also analyzed with varying degrees of exposure uncertainty. The uncertain data with shape parameter $\sigma_1 = 0.5$ are assumed to be real data. On these, further uncertainly with shape parameter σ_2 is added artificially. This is repeated 50 times, and the mean of the results is plotted as a smooth line. The combined uncertainty is calculated using

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}.$$

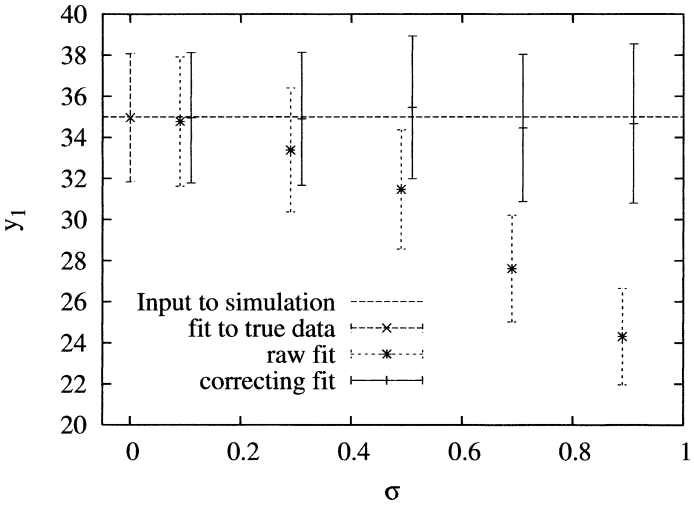

FIG. 5. Effect of Berkson error on the initiation parameter y_1. Shown are the means of the results for 50 simulated data sets with 10,000 records each. The data are analyzed without taking account of the exposure uncertainty (raw fit) and by using the distribution of exposure uncertainty (correcting fit). Neighboring raw and correcting fits are done with the same data. They are slightly shifted along the σ axis to make the presentation clearer.

$$l_w = \frac{1}{N}\sum_{j=1}^{N} l_{x_j}. \tag{8}$$

The sample points x_j remain unchanged during the fit.

RESULTS

Simulation of Additional Error

As a first step, the procedures used for simulation and analysis are tested. Data sets of various sizes are simulated and analyzed by maximizing the likelihood. Depending on the size of the data set and the number of repetitions with different sets of random numbers, the parameters used for the simulation are reproduced reliably by the analysis. A high-quality random number generator is required to avoid bias in large data sets, or many repetitions. The portable code ran1 from the numerical recipes (*23*) fulfills the tests and is used.

Next the effect of exposure uncertainty on the parameters of the TSCE model is studied. As expected, all parameters are affected, but the largest consequences are for the parameters which describe exposure dependence. For the initiating action of exposure, an example simulation is shown in Fig. 4. The fit without uncertainty reproduces the value $y_1 = 35$ from the simulation within the one standard error estimate. When exposure uncertainty is added, the estimated value of y_1 decreases.

A real-life data set with classical error structure and fairly large dose uncertainty might correspond to the point at $\sigma = 0.5$. To simulate a sensitivity analysis for this data set, additional exposure uncertainty is added, and the resulting data are analyzed. This step is repeated 50 times to reduce noise. It can be seen that the effect of applying log-normally distributed uncertainty twice gives about the same result as applying it only once if the shape parameter σ is calculated appropriately. It can also be seen that a sensitivity analysis may give the trend for the effects of the uncertainty correctly.

Note that in this section all data sets used for parameter estimates have the same failure times; they differ only in the recorded exposure rates.

Correction of Berkson Error

Results from including exposure uncertainties with Berkson error structure are given in Figs. 5–8. The average of the maximum likelihood estimate (MLE) and of the standard errors of 50 fits is given. Figure 5 gives the results for the initiation parameter. The fit to the true data reproduces the parameter used in the data simulation well. Adding the log-normally distributed Berkson error does give an attenuation of the parameter y_1, but to a smaller degree than in the classical error situation of Fig. 4. Correcting the fit with the likelihood Eq. (5) reproduces the fit to the true data well. Even for massive distortions of the true exposures for the larger values of σ (compare Fig. 3), the attenuated estimates recover to the correct values. As expected, the estimated error in the parameter estimate increases slightly. Here a different set of failure times is used for each value of uncertainty σ and each of the 50 runs. The raw fit and the corresponding correcting fit are done with the same data.

Comparable results are also obtained when promotion,

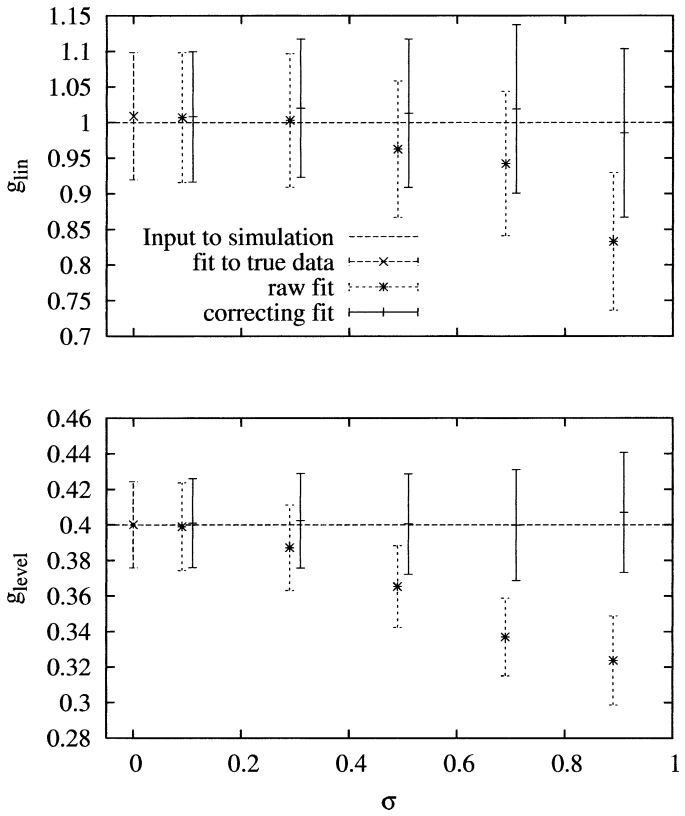

FIG. 6. Effect of Berkson error on the promotion parameters. Shown are the means of the results for 50 simulated data sets with 10,000 records each. The data are analyzed without taking account of the exposure uncertainty (raw fit) and by using the distribution of exposure uncertainty (correcting fit). Neighboring raw and correcting fits are done with the same data. They are slightly shifted along the σ axis to make the presentation clearer.

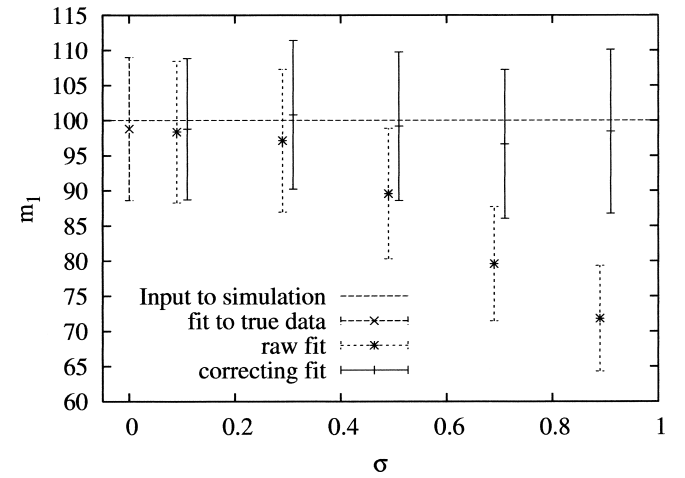

FIG. 7. Effect of Berkson error on the transformation parameter. Shown are the means of the results for 50 simulated data sets with 10,000 records each. The data are analyzed without taking account of the exposure uncertainty (raw fit) and by using the distribution of exposure uncertainty (correcting fit). Neighboring raw and correcting fits are done with the same data. They are slightly shifted along the σ axis to make the presentation clearer.

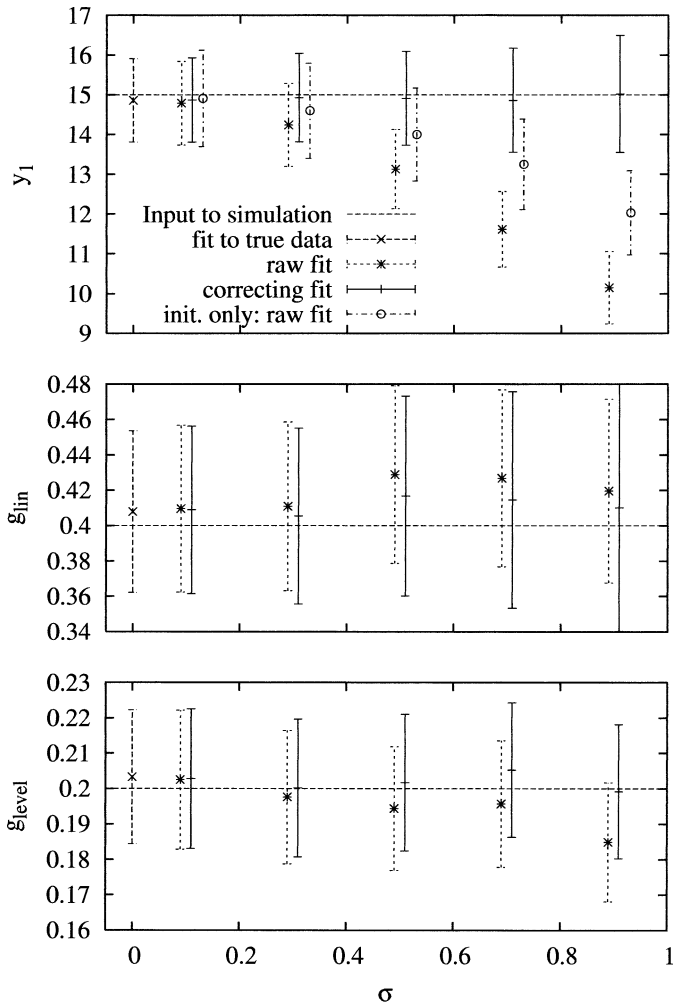

FIG. 8. Effect of Berkson error on the initiation and promotion parameters. Shown are the means of results for 50 simulated data sets with 20,000 records each. The data are analyzed without taking account of the exposure uncertainty (raw fit) and by using the distribution of exposure uncertainty (correcting fit). Neighboring raw and correcting fits are done with the same data. They are slightly shifted along the σ axis to make the presentation clearer. For comparison, results of a simulation and a raw fit are also given when the promotion parameters are fixed to 0.

transformation or combinations of them are made dependent on exposure. Only the effect on the exposure-relevant parameters is shown. Figure 6 gives the result for a promoting action of the exposure, Fig. 7 for a transforming action. In both cases the exposure uncertainty leads to a substantial attenuation of the exposure–response parameters which is corrected by integrating over the distribution of true exposures. As a further example, Fig. 8 shows the result when both initiation and promotion are made dependent on exposure. Here the main effect of dose uncertainty is on the initiation, while the promotion parameters are less affected. To clarify the point, an otherwise equivalent simulation and raw fit was done with the promotion parameters fixed to 0. The attenuation due to dose uncertainty is smaller. No systematic analysis of such shifts in effect was attempted. All parameter estimates are brought back to their

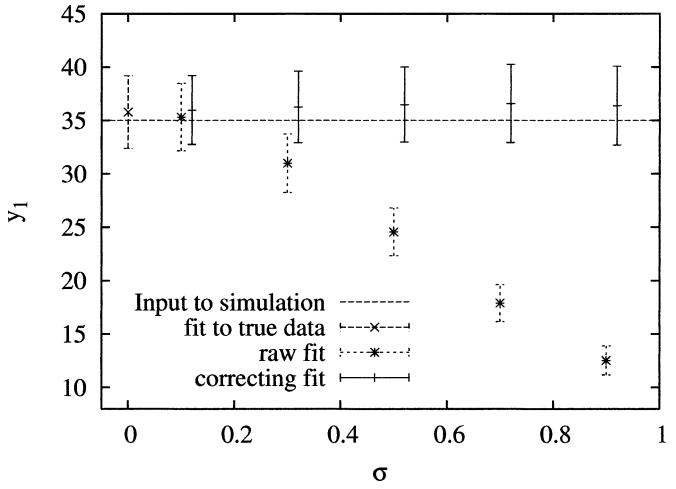

FIG. 9. Effect of classical error on the initiation parameter y_1. Shown are the means of the results for 50 simulated data sets with 10,000 records each. The data are analyzed without taking account of the exposure uncertainty (raw fit) and by using the distribution of exposure uncertainty (correcting fit). Neighboring raw and correcting fits are done with the same data. They are slightly shifted along the σ axis to make the presentation clearer. Note that the corrected values should not converge against the value of y_1 that is used in simulating the data, but against the value estimated by fitting to the true data, since the same set of failure times is used for the different uncertainties in this classical error situation.

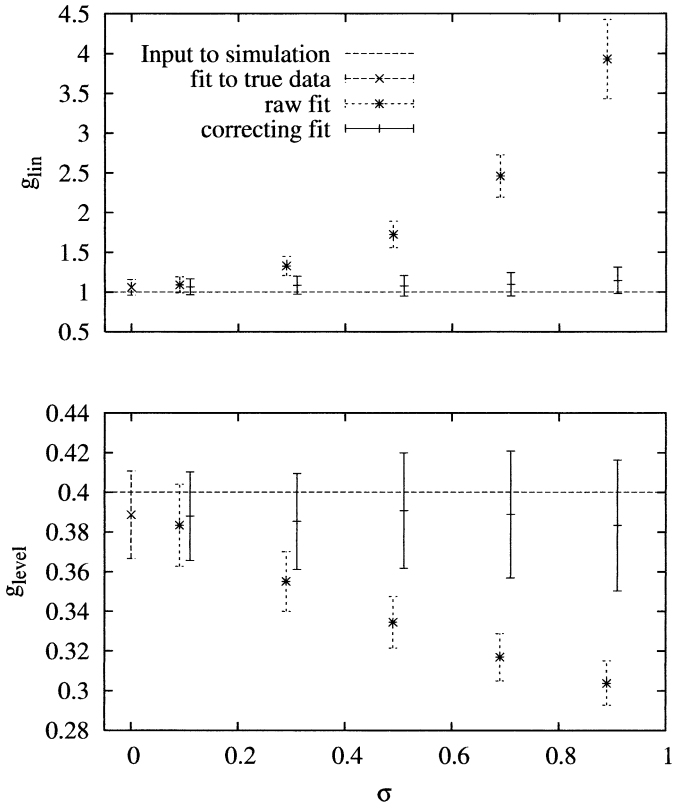

FIG. 10. Effect of classical error on the promotion parameters. Shown are the means of the results for 50 simulated data sets with 10,000 records each. The data are analyzed without taking account of the exposure uncertainty (raw fit) and by using the distribution of exposure uncertainty (correcting fit). Neighboring raw and correcting fits are done with the same data. They are slightly shifted along the σ axis to make the presentation clearer. Note that the corrected values should not converge against the value of y_1 that is used in simulating the data, but against the value estimated by fitting to the true data, since the same set of failure times is used for the different uncertainties in this classical error situation.

true values when the correcting likelihood is used with the correct error distribution.

Correction of Classical Error

Results for including exposure uncertainties with classical error structure are given in Figs. 9–12 for models with the same dose dependence as those used in Figs. 5–8. Again, the average of the MLEs and the average of their standard errors are computed from 50 simulations, with classical error. In all cases the correction mechanism using the correct likelihood (Eq. 6) works very well. However, the effects of the errors in the raw fits can be quite different: Making the initiation parameter dependent on dose (see Fig. 9) gives the same attenuation as in Fig. 4, smoothed by the averaging. For the model with dose-dependent promotion in Fig. 10, the level parameter g_{level} is attenuated, while the linear parameter g_{lin} is amplified. For Berkson errors (compare Fig. 6), both parameters are attenuated in the raw fits. For the transforming action in Fig. 11, the classical errors give a strong amplification in the raw fits, contrary to the situation with Berkson errors in Fig. 7. Here is an example of the fact that the direction of the effect on risk may depend on the error structure (*1*). Finally, in Fig. 12, it can be seen that the effects on the initiation and promotion parameters also occur when both are made dependent on dose. The trends are the same as for Berkson errors (compare Fig. 8) but are more pronounced.

Correcting fits are also done using a deliberately misspecified distribution $p(x)$ for the true exposures in Eq. (6). As expected, a distribution of $p(x)$ that is too wide leads to an under-correction, especially with larger uncertainties σ.

Application to the Colorado Miner Data Set

The earlier analysis of this data set (*15*) suggested an action of radon on initiation and promotion. All nine parameters of the model are newly estimated, including those for the spontaneous hazard and the smoking risk. The emphasis here focuses on the three parameters that are relevant to describe the radon risk. Since the actions of the two parameters which describe promotion are not easily understood numerically, Figs. 13 and 14 are presented. The promoting effect of radon is increased with increasing Berkson uncertainty for both assumed types of errors: For different factors in each interval, the initiating action of radon is slightly decreased; when the same factor is used, it is decreasing strongly, and it would even become negative if no bound at $y_4 = 0$ were used. Thus here is an example of the fact that an exposure effect parameter can be decreased by correcting for exposure uncertainty. But it must be noted

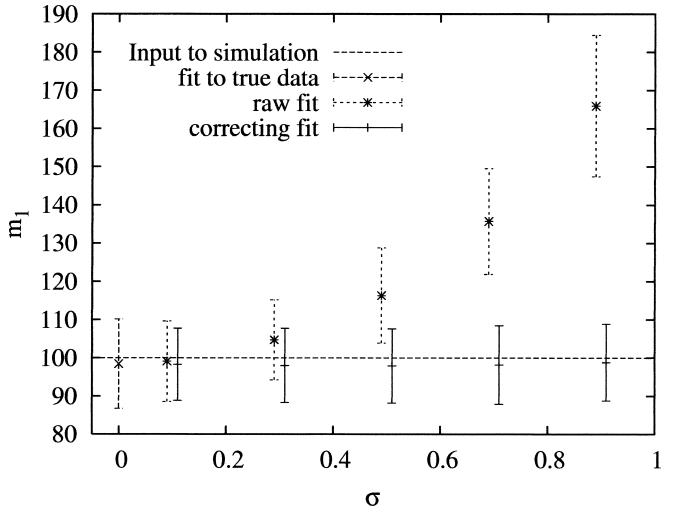

FIG. 11. Effect of classical error on the transformation parameter. Shown are the means of the results for 50 simulated data sets with 10,000 records each. The data are analyzed without taking account of the exposure uncertainty (raw fit) and by using the distribution of exposure uncertainty (correcting fit). Neighboring raw and correcting fits are done with the same data. They are slightly shifted along the σ axis to make the presentation clearer. Note that the corrected values should not converge against the value of y_1 that is used in simulating the data, but against the value estimated by fitting to the true data, since the same set of failure times is used for the different uncertainties in this classical error situation.

that in this model promotion is strongly dominating, and the initiation effect of radiation is minor and is not significant (*15*).

DISCUSSION

Large exposure uncertainty can lead to substantial bias in the estimation of exposure–response parameters in the TSCE model. For initiation, an attenuation due to uncertainties is found. For a transforming action, the direction of the bias depends on the type of error model. When more than a single parameter is affected by exposure, as in promotion or in initiation plus promotion, the situation may be more complicated; both over- and underestimation of parameters is possible. Especially for promotion, uncertainties in the exposure duration may be of great importance, but in most data sets the beginnings and ends of the exposures are well known. Therefore, they are kept fixed in this work.

The likelihood techniques for correcting the effects of exposure uncertainty work very well in all cases if the distribution of error is estimated reliably. This may be more difficult for classical error, when the distribution $p(x)$ of true exposures also has to be estimated. The distribution of exposures could change with time due to the preferential weeding out of the more highly exposed individuals. This fact has not been considered in the current analyses, but the effect is likely to be small, as has been shown by way of simulation.

The likelihood Eq. (8) suggests another, potentially more general approach. The sample points used in the Monte

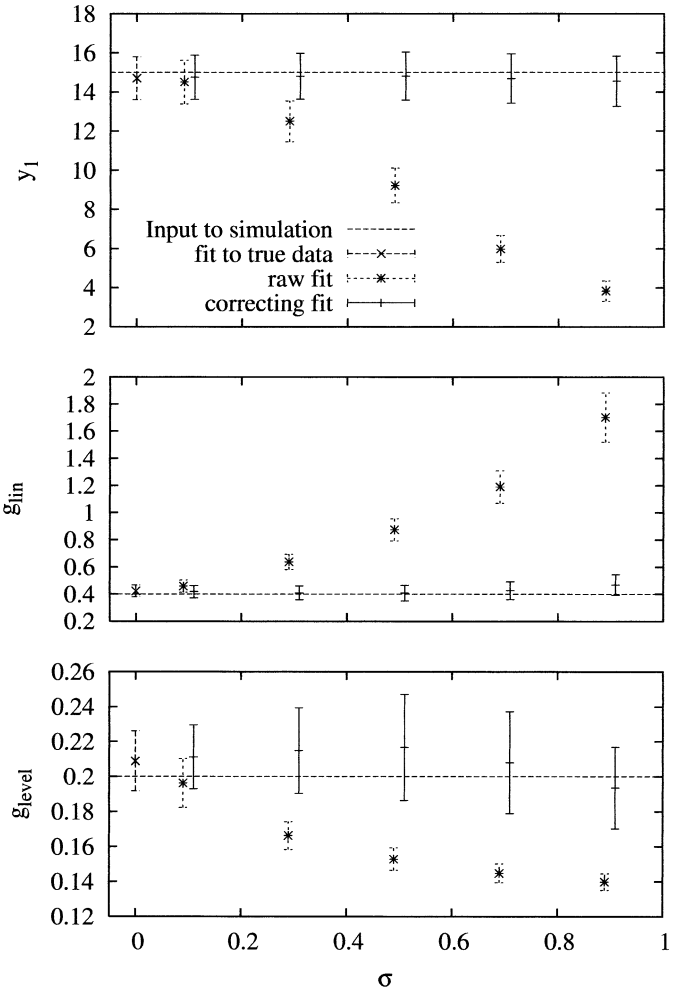

FIG. 12. Effect of classical error on the initiation and promotion parameters. Shown are the means of results for 50 simulated data sets with 20,000 records each. The data are analyzed without taking account of the exposure uncertainty (raw fit) and by using the distribution of exposure uncertainty (correcting fit). Neighboring raw and correcting fits are done with the same data. They are slightly shifted along the σ axis to make the presentation clearer. Note that the corrected values should not converge against the value of y_1 that is used in simulating the data, but against the value estimated by fitting to the true data, since the same set of failure times is used for the different uncertainties in this classical error situation.

Carlo integration are obtained from the independent distributions of exposure uncertainties in various intervals. They are computed from a few numbers which characterize these distributions. If the dosimetrists do not provide a best estimate for exposure and characterize the uncertainty distribution but do provide a sample of say 1000 points from the distribution of estimated true exposures for each person, the risk analysis can be done, taking all this information into account. Such an approach would allow inclusion of correlations, mixtures of random and systematic errors, mixtures of Berkson and classical errors, etc. by the scientists who know the sources of the dose uncertainties best. We are aware that this puts a heavy burden on the dosimetrist, which may not be justified in general. An investi-

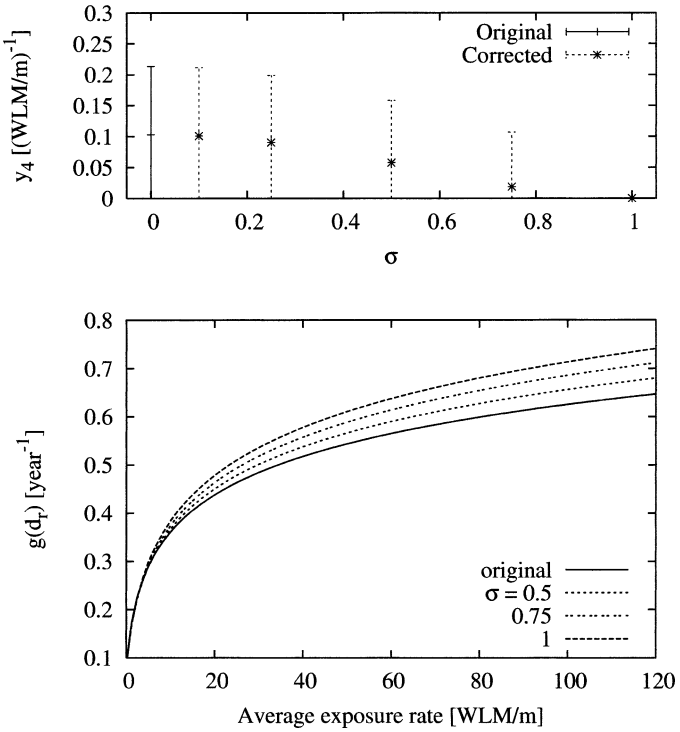

FIG. 13. Colorado miners: Corrected fits if the log-normally distributed uncertainty factors are shared by all radon exposures of a miner. Shown is the effect on the parameter y_4 which describes a linear dependence of initiation on radon, and the effect of the radon rate d_r on promotion $g(dr)$ for a nonsmoker.

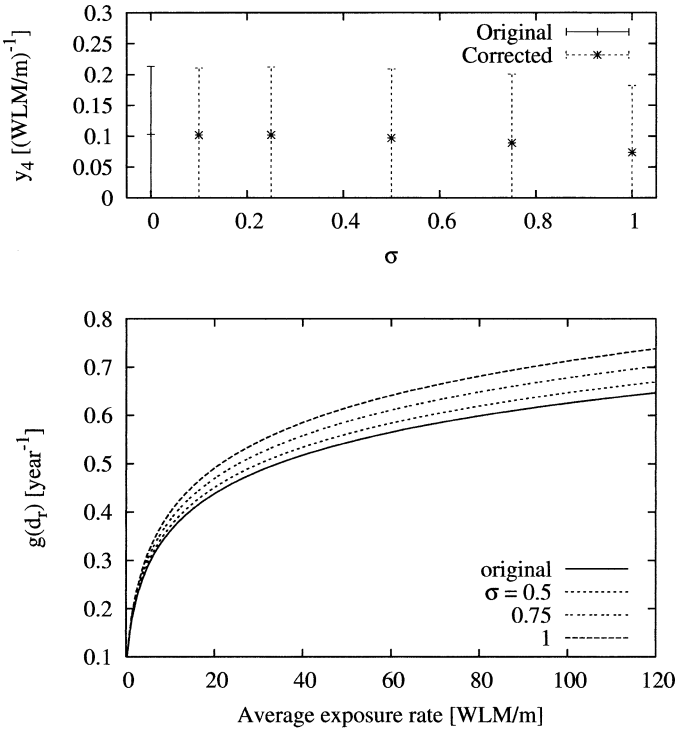

FIG. 14. Corrected fits if different log-normally distributed uncertainty factors are multiplies to each exposure interval of a miner. Shown is the effect on the parameter y_4 which describes a linear dependence of initiation on radon, and the effect of the radon rate d_r on promotion $g(d_r)$ for a nonsmoker.

gation of the full implications of such an approach is beyond the scope of this work.

The promoting effect of radon found in ref. (*15*) comes in part from time-since-exposure effects and in part from protraction effects in the data. Therefore, the fitted model also shows protraction effects. In Fig. 15, Fig. 4 of ref. (*15*) is repeated for two of the exposures, with the effects of exposure uncertainty included. The curves do not change drastically. For the high exposure rate, the risk is increased when correcting for exposure uncertainty. For the low ex-

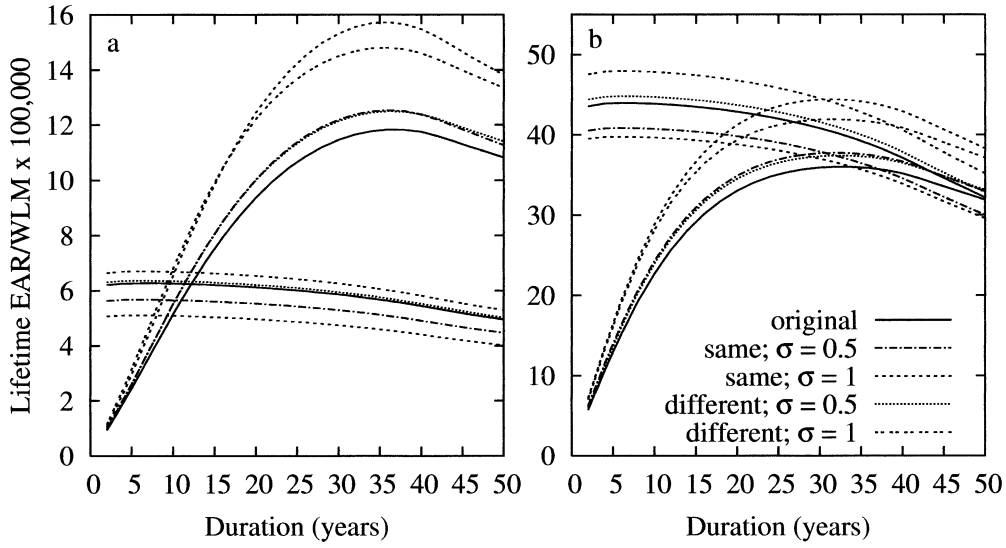

FIG. 15. Lifetime EAR per WLM at age 70 as a function of the duration of exposure (in years) for various total exposures. The exposures are centered at age 40 for (panel a) nonsmokers and (panel b) smokers smoking 10 cigarettes/day starting at age 15. The lines are given for a cumulative exposure to 1 WLM (close to horizontal lines) and 1000 WLM (lines with steep initial increase). The figure corresponds to Fig. 4 of ref. (*15*).

TABLE 1
Comparison of Computed Lifetime Relative Risks (age 70) of Indoor Radon Exposures

Radon exposure		TSCE model				
				$\sigma = 1$		No
WLM/year	Bq/m³	EAD[a]	Original	Same	Different	smoking
Nonsmokers						
0.19	50	1.26	1.06	1.05	1.07	1.075
0.78	200	2.03	1.29	1.25	1.32	1.33
3.12	800	5.01	2.50	2.37	2.73	2.54
Smokers						
0.19	50	1.11	1.05	1.05	1.06	1.075
0.78	200	1.42	1.21	1.21	1.24	1.33
3.12	800	2.51	1.93	1.96	2.08	2.54

Notes. For comparison, estimates from the BEIR VI report (*26*) are given. Also included (in the last column ''no smoking'') is the TSCE result when smoking information in the data set is ignored. The table corresponds to Table 4 of ref. (*15*). The fit without taking exposure uncertainties into account is denoted Original; when the uncertainty factors are shared by all radon exposures of a miner, it is called Same (see also Fig. 13); when they are different for each exposure interval, it is called Different (see also Fig. 14).

[a] Exposure-age-duration model.

posure rate, the trend in risk is dependent on the error model: It is increased when the factors in each interval are different but decreased when they are the same. This general pattern agrees with observations of Stram *et al.* (*7*), where a substantially increased risk due to error corrections was found at high rates but not at low rates.

Another quantity which is given in ref. (*15*) is the extrapolated lifetime relative risk (LRR) for various lifelong exposures to levels of radon as they are found under indoor situations. Table 1 gives the results for the large uncertainty of $\sigma = 1$. When different factors in each interval are used, the estimated LRR increases slightly in all cases. The values for $\sigma = 0.5$ (not shown) lie in between.

No uncertainty analysis was done concerning the smoking information in the data set. But if this information is ignored, and the two smoking-related parameters are left out, then the estimated LRR is only a little higher than that for the nonsmokers in the more detailed version. This indicates some correlation between smoking and radon but not a major one. For the lower exposures, the differences to the exposure-age-duration model of BEIR VI remains substantial.

CONCLUSIONS

The simulation part of this work shows that the likelihood techniques for including exposure uncertainty also work for the complex and highly nonlinear TSCE model. The computational burden can be substantial, maybe a factor of 1000 greater if Monte Carlo integration techniques are necessary, compared to a raw fit. But for the standard version of the TSCE model, it is feasible. Thus these techniques can be put into the toolbox of biologically based modeling with confidence.

If the exposure uncertainties are large and estimable, the analysis should make use of this information; the potential consequences are not easy to predict for the TSCE model, especially if more than one quantity is affected by the exposure. Both an attenuation and an amplification are possible. A cooperation of dosimetrists and risk analysts at an early phase of work might be helpful.

The principal results for the Colorado miners from ref. (*15*) are surprisingly stable when the radon exposure is taken to be uncertain randomly. The estimated risks at high exposure rates are higher. Systematic errors would modify the risk estimates, and a more detailed study would be desirable, taking into account correlations and varying degrees of uncertainties for different exposures. If correction for substantial classical errors are incorporated into the TSCE model, then some reduction in the promoting action of radon is likely, based on the simulations described above. However, our analysis of the Colorado miner data has emphasized Berkson errors, which we believe to be by far the most important component of the exposure measurement errors for that cohort. The hypothesis that radon acts as a promoter survives this test. The mechanism of a promoting action of radon clearly needs further investigation (*24*).

APPENDIX

Simulation of Log-normal Distributions

It is assumed that the uncertainties are multiplicative and are log-normally distributed with a mean of 1 so that no systematic error is produced. The shape parameter σ determines the magnitude of the uncertainty, the scale parameter and the median are $1/e^{\sigma 2/2}$. Log-normally distributed random numbers I_{LN} with mean M are obtained from normally distributed random numbers I_N with zero mean and unit variance by (*25*)

$$I_{LN} = \frac{M}{e^{\sigma^2/2}} e^{\sigma I_N}. \tag{A1}$$

The routine gasdev from the numerical recipes (*23*) is used for calculating normally distributed random numbers.

Integrating over Berkson Errors

When the true exposures x are log-normally distributed, with the mean value at the recorded exposure w then the likelihood Eq. (5) becomes the integral

$$l_w = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty \frac{1}{x} e^{-[\ln(x/w)+\sigma^2/2]^2/(2\sigma^2)} l_x \, dx. \tag{A2}$$

With the transformation

$$u = \frac{\ln(x/w) + \sigma^2/2}{\sqrt{2}\sigma}, \tag{A3}$$

it takes the form

$$l_w = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-u^2} l_{x(u)} \, du. \tag{A4}$$

Gauss-Hermite integration approximates the integral by a sum

$$l_w \approx \frac{1}{\sqrt{\pi}} \sum_i^n v_i l_{x_i}, \tag{A5}$$

where

$$x_i \equiv we^{\sqrt{2}\sigma u_i - \sigma^2/2}.$$

The two n-dimensional vectors u, v are calculated with the routine gauher from the numerical recipes (*23*). Tests showed that $n = 10$ provides sufficient accuracy in this context.

Integrating over Classical Errors

In this case the recorded exposures w are log-normally distributed around the true exposures x, taken as mean value. For a general distribution of true doses $p(x)dx$, the likelihood Eq. (6) takes the form

$$l_w = \frac{1}{c(w)} \frac{1}{\sqrt{2\pi}\sigma w} \int_0^{+\infty} e^{-[\ln(x/w)-\sigma^2/2]^2/(2\sigma^2)} l_x p(x)\, dx. \tag{A6}$$

With the transformation

$$u = \frac{\ln(x/w) - \sigma^2/2}{\sqrt{2}\sigma} \tag{A7}$$

it becomes

$$l_w = \frac{1}{c(w)} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-u^2} l_{x(u)} \frac{x(u)}{w} p[x(u)]\, du. \tag{A8}$$

This can be approximated by Gauss-Hermite integration as

$$l_w \approx \frac{1}{c(w)} \frac{1}{\sqrt{\pi}} \sum_i v_i l_{x_i} \frac{x_i}{w} p(x_i), \tag{A9}$$

where

$$x_i = we^{\sqrt{2}\sigma u_i + \sigma^2/2}.$$

The normalization $c(w)$ can be calculated with the same technique,

$$c(w) \approx \frac{1}{\sqrt{\pi}} \sum_i v_i \frac{x_i}{w} p(x_i). \tag{A10}$$

Combining the two formulas gives the likelihood

$$l_w \approx \frac{\sum_i v_i l_{x_i} x_i p(x_i)}{\sum_i v_i x_i p(x_i)}. \tag{A11}$$

In the simulations, the probability distribution $p(x)dx$ of true exposures is 1 dx for rates between 0 and 1 and is 0 otherwise. Therefore, the sums are taken for those values of x_i, which are smaller than 1. To allow for this finite upper integration limit, 50-dimensional vectors are used.

ACKNOWLEDGMENTS

Most of this work was done while WFH was a visiting scientist at the Fred Hutchinson Cancer Research Center. This work was supported by EU contracts FIGD-CT2000-00083 and FIGH-CT1999-0013 to GSF, and by U.S. Department of Energy contract DE-FC03-99ER62857 and NIH grants R0 1 OH07864 and R0 1 OE509683 to FHCRC.

Received: April 19, 2002; accepted: July 17, 2003

REFERENCES

1. R. J. Carroll, D. Ruppert and L. A. Stefanski, *Measurement Error in Nonlinear Models.* Chapman & Hall, London, 1995.
2. D. Thomas, D. Stram and J. Dwyer, Exposure measurement error: Influence on exposure-disease relationships and methods of correction. *Annu. Rev. Publ. Health* **14**, 69–93 (1993).
3. G. K. Reeves, D. R. Cox, S. C. Darby and E. Whitley, Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Stat. Med.* **17**, 2157–2177 (1998).
4. K. Y. Fung and D. Krewski, On measurement error adjustment methods in Poisson regression. *Environmetrics* **10**, 213–224 (1999).
5. E. S. Gilbert, Accounting for errors in dose estimates used in studies of workers exposed to external radiation. *Health Phys.* **74**, 22–29 (1998).
6. D. A. Pierce, D. O. Stram and M. Vaeth, Allowing for random errors in radiation dose estimates for the atomic bomb survivor data. *Radiat. Res.* **123**, 275–284 (1990).
7. D. O. Stram, B. Langholz, M. Huberman and D. C. Thomas, Correcting for exposure measurement error in a reanalysis of lung cancer mortality for the Colorado plateau uranium miners cohort. *Health Phys.* **77**, 265–275 (1999).
8. F. Lagarde, G. Pershagen, G. Åkerblom, O. Axelson, U. Bäverstam, L. Damber, A. Enflo, M. Svartengren and G. A. Swedjemark, Residential radon and lung cancer in Sweden: Risk analysis accounting for random error in the exposure assessment. *Health Phys.* **72**, 269–276 (1997).
9. *Uncertainties in Radiation Dosimetry and Their Impact on Dose–Response Analysis.* Technical Report No. 99-4541, National Cancer Institute, Bethesda, MD, 1999.
10. A. G. Knudson, Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* **68**, 820–823 (1971).
11. S. H. Moolgavkar and G. Luebeck, Two-event model for carcinogenesis: Biological, mathematical, and statistical considerations. *Risk Anal.* **10**, 323–341 (1990).
12. W. F. Heidenreich, E. G. Luebeck and S. H. Moolgavkar, Some properties of the hazard function of the two-mutation clonal expansion model. *Risk Anal.* **17**, 391–399 (1997).
13. W. F. Heidenreich, P. Jacob and H. G. Paretzke, Solutions of the clonal expansion model and their application to the tumor incidence of the atomic bomb survivors. *Radiat. Environ. Biophys.* **36**, 45–58 (1997).
14. M. Kai, E. G. Luebeck and S. H. Moolgavkar, Analysis of the incidence of solid cancer among atomic bomb survivors using a two-stage model of carcinogenesis. *Radiat. Res.* **148**, 348–358 (1997).
15. E. G. Luebeck, W. F. Heidenreich, W. D. Hazelton, H. G. Paretzke and S. H. Moolgavkar, Biologically based analysis of the Colorado uranium miners cohort data: Age, dose and dose-rate effects. *Radiat. Res.* **152**, 339–351 (1999).
16. E. G. Luebeck, S. B. Curtis, F. T. Cross and S. H. Moolgavkar, Two-stage model of radon-induced malignant lung tumors in rats: Effects of cell killing. *Radiat. Res.* **145**, 163–173 (1996).
17. W. F. Heidenreich, P. Jacob, H. G. Paretzke, F. T. Cross and G. E. Dagle, Two-step model for fatal and incidental lung tumor risk in rats exposed to radon. *Radiat. Res.* **151**, 209–217 (1999).
18. W. F. Heidenreich, M. J. P. Brugmans, M. P. Little, H. P. Leenhouts, H. G. Paretzke, M. Morin and J. Lafuma, Analysis of lung tumour risk in radon-exposed rats: An intercomparison of multi-step modelling. *Radiat. Environ. Biophys.* **39**, 253–264 (2000).
19. W. F. Heidenreich, On the parameters of the clonal expansion model. *Radiat. Environ. Biophys.* **35**, 127–129 (1996).
20. F. James, *Minuit Function Minimization and Error Analysis, version 94.1.* CERN, Geneva, 1994.
21. V. E. Archer, J. K. Wagoner and F. E. Lundin, Lung cancer among uranium miners in the United States. *Health Phys.* **25**, 351–371 (1973).
22. W. H. Press, B. P. Flannery, S. A. Teukolski and W. T. Vetterling, *Numerical Recipes.* Cambridge University Press, Cambridge, 1992.
23. W. H. Press, S. A. Teukolski, W. T. Betterling and B. P. Flannery, *Numerical Recipes in FORTRAN*, 2nd ed. Cambridge University Press, Cambridge, 1992.
24. W. F. Heidenreich, M. Atkinson and H. G. Paretzke, Radiation-induced cell inactivation can increase the cancer risk. *Radiat. Res.* **155**, 870–872 (2001).
25. N. A. Hastings and J. B. Peacock, *Statistical Distributions.* Wiley, New York, 1975.
26. National Research Council, Committee on Health Risks of Exposure to Radon, *Health Effects of Exposure to Radon (BEIR VI).* National Academy Press, Washington, DC, 1999.