Methods for Estimation of Radiation Risk in Epidemiological Studies Accounting for Classical and Berkson Errors in Doses

Alexander Kukush, *Ukraine Radiation Protection Institute*
Sergiy Shklyar, *Ukraine Radiation Protection Institute*
Sergii Masiuk, *Ukraine Radiation Protection Institute*
Illya Likhtarov, *Ukraine Radiation Protection Institute*
Lina Kovgan, *Ukraine Radiation Protection Institute*
Raymond J. Carroll, *Texas A&M University*
Andre Bouville, *National Cancer Institute, NIH, DHHS*

Methods for Estimation of Radiation Risk in Epidemiological Studies Accounting for Classical and Berkson Errors in Doses

Alexander Kukush, Sergiy Shklyar, Sergii Masiuk, Illya Likhtarov, Lina Kovgan, Raymond J. Carroll, and Andre Bouville

Abstract

With a binary response Y, the dose-response model under consideration is logistic in flavor with $pr(Y=1 \mid D) = R (1+R)^{-1}$, $R = \lambda_0 + \text{EAR } D$, where λ_0 is the baseline incidence rate and EAR is the excess absolute risk per gray. The calculated thyroid dose of a person i is expressed as $D_i^{mes} = f_i Q_i^{mes}/M_i^{mes}$. Here, Q_i^{mes} is the measured content of radioiodine in the thyroid gland of person i at time t_{mes}, M_i^{mes} is the estimate of the thyroid mass, and f_i is the normalizing multiplier. The Q_i and M_i are measured with multiplicative errors V_i^Q and V_i^M, so that $Q_i^{mes} = Q_i^{tr} V_i^Q$ (this is *classical* measurement error model) and $M_i^{tr} = M_i^{mes} V_i^M$ (this is *Berkson* measurement error model). Here, Q_i^{tr} is the true content of radioactivity in the thyroid gland, and M_i^{tr} is the true value of the thyroid mass. The error in f_i is much smaller than the errors in (Q_i^{mes}, M_i^{mes}) and ignored in the analysis.

By means of Parametric Full Maximum Likelihood and Regression Calibration (under the assumption that the data set of true doses has lognormal distribution), Nonparametric Full Maximum Likelihood, Nonparametric Regression Calibration, and by properly tuned SIMEX method we study the influence of measurement errors in thyroid dose on the estimates of λ_0 and EAR. The simulation study is presented based on a real sample from the epidemiological studies. The doses were reconstructed in the framework of the Ukrainian-American project on the investigation of Post-Chernobyl thyroid cancers in Ukraine, and the underlying subpolulation was artificially enlarged in order to increase the statistical power. The true risk parameters were given by the values to earlier epidemiological studies, and then the binary response was simulated according to the dose-response model.

1   Introduction

The current methods of risk estimation (see *Health Risks from Exposure to Low Levels of Ionizing Radiation*, 2006; Jacob et al., 2006; Likhtarev et al., 2006a) take into account uncertainties related to health effects, but assume that the doses are known with perfect precision and accuracy, something that is not true (Kopecky et al., 2006; Likhtarev et al., 1993; Likhtarev et al., 2003; Ron and Hoffman, 1999). Although many attempts have been made to develop mathematical tools which take into account the dose uncertainty (see Carroll et al., 2006; Hofer, 2008; Kopecky et al., 2006; Li et al., 2007; Masiuk et al., 2008), the problem has yet to be solved in a satisfactory manner. It is our opinion that one of the reasons an accurate model is so hard to create is because the dose estimates are almost inevitably associated with a mixture of classical and Berkson errors (Reeves et al., 1998; Carroll et al., 2006; Li et al., 2007), which leads to difficulties in evaluating the influence of the mixed errors in the estimated dose on risk analysis, which is usually expressed in terms of relative (ERR) or absolute (EAR) risk (see *Health Risks from Exposure to Low Levels of Ionizing Radiation*, 2006). One such example is the analysis of epidemiological investigations of thyroid cancer among children exposed to radioiodines in fallout from the Chornobyl accident (Jacob et al., 2006; Likhtarev et al., 2006a; Tronko et al., 2006).

In the Chornobyl accident, not only the absolute and relative frequencies of thyroid cancer were determined with accuracy, but also the thyroid doses and their uncertainties of Berkson and classical types were carefully estimated. However, computationally efficient procedures that account for the mixed nature of the dosimetric uncertainties in the risk analysis are not available.

In the epidemiologic study of thyroid cancer conducted in Ukraine (Likhtarev et al., 2006b), the thyroid dose estimates are based on the measurements of Iodine-131 (^{131}I) that were performed on all subjects, and the uncertainties in the thyroid dose estimates are mainly due to those in the measured activities and in the thyroid mass (Likhtarev et al., 2003).

It is assumed in this paper that the thyroid mass is measured with Berkson error, which appears as a result of averaging, while uncertainty in the measured activity is classical. The presence of errors of both types leads to a mixture of classical and Berkson errors in the thyroid dose estimates. In this case, the two components of the uncertainty in the dose are quite distinct and quantifiable, unlike in Reeves et al. (1998) and Li et al. (2007), where the mixture error model involves, besides the true dose, additional, unobservable variables in the model of dose measurement.

The model proposed in this paper is a new way of thinking about these problems. In the Nevada Test Site Study (Apanasovich et al., 2009; Li et al., 2007;

Lyon et al., 2006; Mallick et al., 2002) it was recognized that there was a mixture of Berkson and classical errors in the dose estimates, but the available data set did not have information on their relative sizes, and hence the statistical methods used here are quite different. Even the SIMEX version in Apanasovich et al. (2009) is different from the modification developed below because of the different dose uncertainty model.

Though in the simulations and analysis of the source of uncertainty induced by Berkson and classical measurement errors we used the dose estimate and its uncertainty related to the thyroid exposure (Likhtarev et al., 1993; Likhtarev et al., 2003; Likhtarev et al., 2005), the main goal of the present paper is to study the possibility to apply in the risk analysis the most popular methods of regression analysis, e.g., Simulation-Extrapolation (SIMEX), Maximum Likelihood and Regression Calibration (see Carroll et al., 2006; Kopecky et al., 2006; Masiuk et al., 2008; Kukush and Schneeweiss, 2005). Those methods take into account the presence of measurement errors (classical, Berkson and mixed) in regressors. In particular we study the reliability of radiation risk estimates obtained by each of those methods, as well as their robustness with respect to the error magnitude and the form of distribution of the initial data set of dose estimates. Attention will be paid to development and mathematical justification of modifications of the listed methods that under certain conditions lead to significant improvement of the estimates.

In order to get the results that are as much as possible independent of the concrete radio-epidemiological investigations, a cohort was simulated with the following properties:

- The cohort is sufficiently large that there are enough thyroid cancer disease cases for the existence of a solution to the optimization of the objective log-likelihood function, for all simulated data sets. Thyroid cancer is rare in a population, even for the Chornobyl accident.
- The radiation risk parameters used are close to the quantities obtained in Jacob, et al. (2006), Likhtarev, et al. (2006a) and Tronko, et al. (2006).
- The risk model is the simplest but quite popular generalized linear model of absolute risk with binary response (see below) that makes it possible to verify and compare the results with estimates obtained by the well-known program package *EPICURE*.

We mention the main novelties of the present paper:

- An efficient version of SIMEX that provides smaller estimation bias compared with ordinary SIMEX.

- A refined version of SIMEX in equation (A.3) from Appendix that improves statistical performance.
- A new nonparametric regression calibration procedure.
- A new nonparametric full maximum likelihood estimating procedure.
- A simulation study in a unique context.

2 Error-Free Dose-Response Model and Risk Estimation

In this section, we describe the error-free model and computation methods for fitting it.

2.1 The Dose-Response Model

With a binary response Y, our dose-response model is

$$\text{pr}(Y = 1|D) = \frac{\lambda_0(1 + \beta D)}{1 + \lambda_0(1 + \beta D)} \tag{1}$$

where D denotes the individual thyroid dose for a given person, the event $Y = 1$ means that the person developed thyroid cancer within a certain time interval (e.g., during ten years), λ_0 is the baseline incidence rate, and $\lambda_0\beta$ is the excess absolute risk per gray (EAR). Of course, λ_0 and β are both positive.

2.2 Maximum Likelihood

The parameters in model (1) can be estimated by maximum likelihood, which consists of solving the equations

$$\begin{array}{rcl}
\sum_{i=1}^{N} Y_i\{\lambda_0(1 + \beta D_i)\}^{-1} & = & \sum_{i=1}^{N}\{1 + \lambda_0(1 + \beta D_i)\}^{-1}; \\
\sum_{i=1}^{N} Y_i D_i\{\lambda_0(1 + \beta D_i)\}^{-1} & = & \sum_{i=1}^{N} D_i\{1 + \lambda_0(1 + \beta D_i)\}^{-1}.
\end{array}$$

These equations can be obtained by differentiating the loglikelihood function (see e.g. Masiuk, 2008, page 67).

<u>**Remark**</u> **1** Of course, standard programs such as *EPICURE* can be used to compute the maximum likelihood estimators, but when we account for measurement error, one of the methods we discuss is the SIMEX algorithm, which is computationally expensive and involves fitting the maximum likelihood method many times.

3

If the bootstrap is employed for inference, the maximum likelihood estimator would have to be computed tens of thousands of times. For this reason, in Section 2.3 we will suggest another approach.

2.3 A Computationally Efficient Method

In order to obtain a computationally efficient estimator, we first note that $\mathbf{E}\{(1-Y)(1+\beta D)\} = \mathbf{E}(Y/\lambda_0)$ and $\mathbf{E}\{(1-Y)\} = \mathbf{E}[Y/\{\lambda_0(1+\beta D)\}]$. Replacing the expectations by their empirical means, this suggests solving

$$0 = \sum_{i=1}^{N} \{(1-Y_i)(1+\beta D_i) - Y_i/\lambda_0\}; \tag{2}$$

$$0 = \sum_{i=1}^{N} [1-Y_i - Y_i/\{\lambda_0(1+\beta D_i)\}]. \tag{3}$$

From (2), for any β we have

$$\widehat{\lambda}_{0Q} = \widehat{\lambda}_{0Q}(\beta) = \sum_{i=1}^{N} Y_i / \{\sum_{i=1}^{N}(1-Y_i)(1+\beta D_i)\}.$$

From (2) and (3), by some algebra, we get the equation for $\widehat{\beta}_Q$ as

$$0 = \sum_{i=1}^{N} Y_i(D_i - D_{\mathrm{av}})/(1+\beta D_i), \quad \beta > 0, \tag{4}$$

where

$$D_{\mathrm{av}} = \sum_{i=1}^{N} D_i(1-Y_i)/\{\sum_{i=1}^{N}(1-Y_i)\}. \tag{5}$$

The statistical and large sample properties of the solution to (4) are described in Appendix A.1.

The solution to (4) is readily calculated via any one-dimensional equation solving method. While not the maximum likelihood estimator, in simulations not shown here, we have found that it is nearly as efficient, and in any case can be used as an initial estimate. When we later wish to employ the SIMEX procedure for measurement error correcting, this computational efficiency becomes important, see Remark 1.

3 Model for Dose

According to the dosimetric model described by Likhtarev et al. (2005, 2006b), the calculated thyroid dose of person i is expressed as

$$D_i^{ins} = \frac{Q_i^{mes}}{M_i^{est}} f_i, \tag{6}$$

where Q_i^{mes} is the measured content of ^{131}I in the thyroid gland of person i at time t_{mes}, M_i^{est} is the estimate of the thyroid mass, and f_i is a multiplier which takes into

4

account the parameter values of an ecology-metabolism model for person i. While f_i is also measured with error, its error is much smaller than the errors in Q_i^{mes} and in M_i^{est} (see Likhtarev et al., 2003, 2006b) and will be ignored in this analysis. In principle, the uncertainty in f_i can be included in the evaluation in many ways, using for example the shared parameter methods of Stram and Kopecky (2003).

The uncertainties in Q_i^{mes} and M_i^{est} are as follows:

- The measured activity is associated with a multiplicative error, which is determined by the characteristics of the measuring instrument (Likhtarev, et al., 1993; Likhtarev, et al., 2003), so that

$$Q_i^{mes} = Q_i^{tr} V_i^Q, \tag{7}$$

where Q_i^{tr} is the true ^{131}I content in the thyroid gland, and V_i^Q is the independent multiplicative measurement error. Expression (7) of course defines a classical multiplicative error model.

- The true values of the thyroid mass M_i^{tr} are determined according to a Berkson measurement error model as

$$M_i^{tr} = M_i^{mes} V_i^M, \tag{8}$$

where M_i^{mes} is the median of the thyroid mass for a given gender-age group, and V_i^M an independent multiplicative error.

- The measurement errors (V_i^Q, V_i^M) are assumed to be lognormally distributed, so that $\log(V_i^Q) = \text{Normal}(0, \sigma_{Q,i}^2)$ and $\log(V_i^M) = \text{Normal}(0, \sigma_{M,i}^2)$. In the Chornobyl data, the parameters $\sigma_{Q,i}^2$ and $\sigma_{M,i}^2$ have been reliably estimated (see Likhtarev et al., 1993), therefore in this paper those variables are assumed known.

- We assume that the random vector (Q_i^{tr}, M_i^{mes}) and random variables V_i^Q, V_i^M are jointly independent, and we set $M_i^{est} = M_i^{mes}$. We allow the random variables Q_i^{tr} and M_i^{mes} to be correlated.[1]

The unobservable error-free dose is denoted by D_i^{tr}, and is given as

$$D_i^{tr} = Q_i^{tr} f_i / M_i^{tr}. \tag{9}$$

---

[1]For the subpopulation from Section 6.1 the sample correlation between $\log(Q^{tr})$ and $\log(M^{mes})$ is equal to 0.26. The explanation for positive correlation between the thyroid activity and the thyroid gland mass is the following: the larger age of a person the larger the mass, and therefore, in general, the larger the activity.

5

Remark 2 The assumption that the measurement errors in the thyroid mass have lognormal distributions is justified as follows: our research group possesses the data set of thyroid volume measurements performed by the Sasakawa Foundation in Ukraine for each age-sex-region group (the quantiles of the corresponding distributions are published in *Chernobyl: A Decade*, 1997, pp. 393–398). It is clear from the data set that the distributions are close to lognormal. By physical reasoning the measured activity is non-negative. Therefore, the multiplicative error in the activity is modeled by a lognormal law. In future we will study the distribution of this error in more detail.

4 Regression Calibration

4.1 Parametric Calibration

In practice, regression calibration (see Carroll et al., 2006) is often used to take into account the errors in measured dose. According to this method, prior to statistical processing of the epidemiological data, the measured dose is replaced by the conditional expectation of the true dose given measured quantities. In our case, we the measured dose is replaced by

$$D_i^* = \mathbf{E}(D_i^{tr} \mid D_i^{ins}), \tag{10}$$

which is really a short-hand notation for conditioning on the measured quantities $(M_i^{est}, Q_i^{mes}, f_i)$. Following this, classical regression analysis of the epidemiological data is performed, e.g., by the program package *EPICURE* (Preston et al., 1993), with inference using a bootstrap.

Consider for example the case that fQ^{tr}/Mmes has a lognormal distribution. Let $\log(fQ^{tr}/M^{mes}) = \text{Normal}(\mu_1, \sigma_1^2)$. Because $\log(D_i^{ins}) = \log(f_i Q_i^{tr}/M_i^{mes}) + \log V_i^Q$ and the summands are independent, the conditional distribution of $\log(f_i Q_i^{tr}/M_i^{mes})$ given D_i^{ins} is normal,

$$\log\left(\frac{f_i Q_i^{tr}}{M_i^{mes}}\right) \Big| D_i^{ins} = \text{Normal}\left(\frac{\sigma_1^2 \log D_i^{ins} + \sigma_{Q,i}^2 \mu_1}{\sigma_1^2 + \sigma_{Q,i}^2}, \frac{\sigma_1^2 \sigma_{Q,i}^2}{\sigma_1^2 + \sigma_{Q,i}^2}\right),$$

see the corresponding formula in Anderson (1958, Section 2.5). Then

$$\mathbf{E}\left(\frac{f_i Q_i^{tr}}{M_i^{mes}} \Big| D_i^{ins}\right) = \exp\left(\frac{\sigma_1^2 \log D_i^{ins} + \sigma_{Q,i}^2 \mu_1 + \sigma_1^2 \sigma_{Q,i}^2/2}{\sigma_1^2 + \sigma_{Q,i}^2}\right).$$

6

Because $D_i^{tr} = (f_i Q_i^{tr}/M_i^{mes}) \times (V_i^M)^{-1}$, and the factors $f_i Q_i^{tr}/M_i^{mes}$ and $(V_i^M)^{-1}$ are independent,

$$
\begin{aligned}
\mathbf{E}[D_i^{tr} \mid D_i^{ins}] &= \mathbf{E}[f_i Q_i^{tr}/M_i^{mes} \mid D_i^{ins}] \mathbf{E}(V_i^M)^{-1} \\
&= \exp \left\{ \frac{\sigma_1^2 \log D_i^{ins} + \sigma_{Q,i}^2 \mu_1 + \sigma_1^2 \sigma_{Q,i}^2/2}{\sigma_1^2 + \sigma_{Q,i}^2} + \frac{\sigma_{M,i}^2}{2} \right\}.
\end{aligned}
$$

Finally, we have to estimate μ_1 and σ_1^2 and substitute the estimates:

$$
D_i^* = \exp \left\{ \frac{\widehat{\sigma}_1^2 \log D_i^{ins} + \sigma_{Q,i}^2 \widehat{\mu}_1 + \widehat{\sigma}_1^2 \sigma_{Q,i}^2/2}{\widehat{\sigma}_1^2 + \sigma_{Q,i}^2} + \frac{\sigma_{M,i}^2}{2} \right\}. \tag{11}
$$

Consistent estimates for μ_1 and σ_1^2 are the sample mean of $\log(D_i^{ins})$ and the sample variance of $\log(D_i^{ins})$ minus the average value of $\sigma_{Q,i}^2$.

Also, the calibration of the dose can be done under the assumption that the distribution of $\log(f Q^{tr}/M^{mes})$ is a mixture of normal distributions. But in our simulations we do not see the advantage of involving the mixture of normals. The reason for that is the unimodality of the distribution of $\log(f Q^{tr}/M^{mes})$, see Figures 1 and 2 in Section 6.2.

4.2   Nonparametric Regression Calibration

Now, we suppose that the form of distribution of $f Q^{tr}/M^{mes}$ is not specified. One way to handle this problem is via a discrete approximation to this distribution. Let the range of $\log(f Q^{tr}/M^{mes})$ be partitioned into K intervals, with midpoints $x_1, \ldots, x_K$. Then construct a discrete approximation to the distribution of $f_i Q_i^{tr}/M_i^{mes}$ according to the formula

$$
\mathrm{pr}\{f_i Q_i^{tr}/M_i^{mes} = \exp(x_k)\} = b_k, \qquad k = 1, \ldots, K \tag{12}
$$

(in the simulation study we took $K = 25$). Here the values b_k are estimated by maximum likelihood based on known distribution of the measurement errors V_i^Q and using the convex optimization technique. Next, by Bayes' formula we construct the calibrated doses

$$
D_i^* = \frac{\sum_{k=1}^K b_k \ell_{ik} \exp(x_k + \sigma_{M,i}^2/2)}{\sum_{k=1}^K b_k \ell_{ik}},
$$

where

$$
\ell_{ik} = \frac{1}{\sqrt{2\pi\sigma_{Q,i}^2}} \exp \left(-\frac{(x_k - \log D_i^{mes})^2}{2\sigma_{Q,i}^2} \right)
$$

is the conditional density of $\log(D_i^{mes})$ given $f_i Q_i^{tr}/M_i^{mes}$ at point x_k.

7

5 The SIMEX Estimator and Its Computationally Efficient Modification

The estimator obtained by the SIMEX method is randomized, i.e., it is a random function of observations. A similar method of risk parameters estimation was used in Kopecky et al. (2006), but the approach proposed below makes it possible to give proper weight to the structure of measured doses. Note that the SIMEX method does not require the knowledge of the probability distribution for the true activity Q_i^{tr}.

To account for the Berkson error, the thyroid mass M_i^{mes} is calibrated by formula:

$$\mathbf{E}\{(1/M_i^{tr})|M_i^{mes}\} = \mathbf{E}(1/V_i^M)/M_i^{mes} = \exp(\sigma_{M,i}^2/2)/M_i^{mes}. \qquad (13)$$

We intend to substitute $\mathbf{E}\{(1/M_i^{tr})|M_i^{mes}\}$ for $1/M_i^{mes}$. In order to take the classical error into consideration, the following steps are taken:

1. Select a natural number $B \geq 2$. Generally, we select B to be a fairly large number, e.g., $B = 100$. Simulate random perturbations of log-activities $U_{b,i}^* = \mathrm{Normal}(0, \sigma_{Q,i}^2)$ for $b = 1, ..., B$ and $i = 1, ..., N$.
2. Choose various levels of $\kappa \geq 0$. We use $\kappa = (0.0, 0.2, 0.4, 0.6)$.
3. For each κ, obtain the perturbed doses

$$D_{b,i}^*(\kappa) = \frac{Q_i^{mes}}{M_i^{mes}} f_i \exp\left(\sigma_{M,i}^2/2 + \sqrt{\kappa} U_{b,i}^*\right). \qquad (14)$$

4. Compute the ordinary (naïve) estimators $\widehat{\lambda}_{0,b}^*(\kappa), \widehat{\beta}_b^*(\kappa)$ for $\kappa = 0$, 0.2, 0.4, 0.6 and average them across b, defining $\widehat{\lambda}_0^*(\kappa) = B^{-1}\sum_{b=1}^B \widehat{\lambda}_{0,b}^*(\kappa)$, $\mathrm{EAR}^*(\kappa) = B^{-1}\sum_{b=1}^B \widehat{\lambda}_{0,b}^*(\kappa)\widehat{\beta}_b^*(\kappa)$.
5. Extrapolate numerically the functions $\widehat{\lambda}_0^*(\kappa)$ and $\mathrm{EAR}^*(\kappa)$ to the point $\kappa = -1$ and get the SIMEX estimator for the parameters λ_0 and $\mathrm{EAR} = \lambda_0\beta$.

We describe in Remark 1 why a computationally efficient version of SIMEX is desirable in our context. In the computationally efficient modification of the SIMEX method, other values are used instead of the doses (14), namely

$$D_{b,i}^*(\kappa) = \begin{cases} (Q_i^{mes}/M_i^{mes})f_i\exp\left\{\left(\sigma_{M,i}^2 - \sigma_{Q,i}^2\right)/2\right\}, & Y_i = 0, \\ (Q_i^{mes}/M_i^{mes})f_i\exp\left(\sigma_{M,i}^2/2 + \sqrt{\kappa}U_{b,i}^*\right), & Y_i = 1, \end{cases}$$

and as a result the system of equations for the modified estimate takes the form

$$\sum_{i=1}^{N}(1-Y_i)\left[1+\widehat{\beta}_b^*(\kappa)\frac{Q_i^{mes}}{M_i^{mes}}f_i\exp\{(1/2)(\sigma_{M,i}^2-\sigma_{Q,i}^2)\}\right]=\sum_{i=1}^{N}Y_i/\widehat{\lambda}_{0,b}^*(\kappa),$$

(15)

$$\sum_{i=1}^{N}(1-Y_i)=\sum_{i=1}^{N}Y_i[\widehat{\lambda}_{0,b}^*(\kappa)\{1+\widehat{\beta}_b^*(\kappa)D_{b,i}^*(\kappa)\}]^{-1}.$$

(16)

Hence the left-hand side of (15) is obtained after taking into account the dose errors and correction of the expression $\sum_{i=1}^{N}(1-Y_i)(1+\beta D_i)$, see equation (2) and Carroll et al. (2006), while the right-hand side of equation (16) is obtained after perturbation of D_i^{Mcal} in the expressions $\sum_{i=1}^{N}Y_i/\{\lambda_0(1+\beta D_i^{Mcal})\}$, where $D_i^{Mcal}=(Q_i^{mes}/M_i^{mes})f_i\exp(\sigma_{M,i}^2/2)$ is the dose calibrated with respect to the errors in the thyroid mass. Equations (15)–(16) can be solved similarly to equations (2)–(3).

This modification of SIMEX is computationally efficient for the following reasons:

- Each evaluation of the naïve estimator, that is solving (15)–(16), is efficient.
- Some calculations common to different evaluations of the naïve estimator have to be performed only once.

A more detailed description of the proposed procedure is presented in Appendix A.2, while the characteristics of its implementation are given in Appendix A.3. This modification of the SIMEX estimator is justified in Appendix A.2.2.

6   Simulation

6.1   Simulation Setup

We performed a simulation study based on data from epidemiological investigations (see Jacob, et al., 2006; Likhtarev, et al., 2005; Likhtarev, et al., 2006a; Likhtarev, et al., 2006b). For the simulation of doses a real subpopulation of 7077 children and teenagers up to 18 years was used; these persons resided in settlements of Zhitomir, Kyiv, and Chernigov Oblasts of Ukraine at the time of the Chornobyl accident and had direct measurements of thyroid activity in May-June 1986. The thyroid doses for this subpopulation were reconstructed in the framework of the Ukrainian-American project on the investigation of Post-Chornobyl thyroid cancers in Ukraine (Likhtarev, et al., 2006b).

6.2   Two Simulation Scenarios

We simulated under two scenarios, in one of which the distribution of thyroid doses is well-modeled by a lognormal distribution, while in the other it is not.

- In the *first scenario* the subpopulation was formed in such a way that the persons with dose fQ^{tr}/Mmes greater than 2 Gray were deleted. In addition, in order to increase the power of the numerical experiment, this subpopulation was artificially enlarged to $N = 66\,665$ persons. The cutoff was performed only at the preliminary stage of the simulation. But after the simulation the doses were not cut off and varied from 3×10^{-7} to 10.5 Grays (see Figure 1).
- In the *second scenario*, at the preliminary stage the cutoff was not performed, and the subpopulation was artificially enlarged to $N = 70\,770$ persons (see Figure 2).
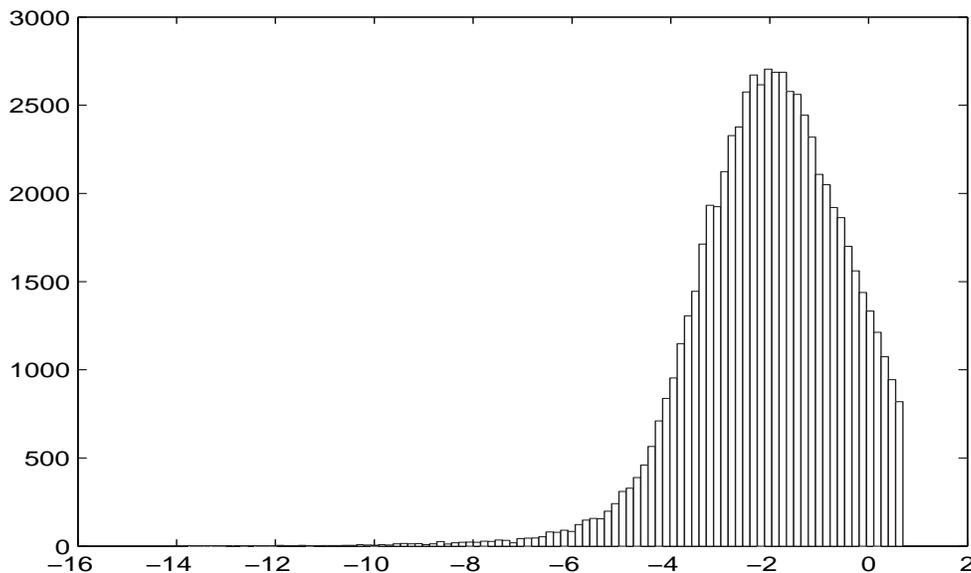
Figure 1: Censored distribution of $\log(fQ^{tr}/M^{mes})$.

The true parameters of the absolute risk model (1) were given by the values close to the estimates obtained in the epidemiological studies of the thyroid cancers in Ukraine (Jacob, et al., 2006; Likhtarev, et al., 2006a). Namely, we set

$$\lambda_0 = 10^{-3}\,\frac{\text{cases}}{\text{persons}}; \qquad EAR = 5 \times 10^{-3}\,\frac{\text{cases}}{\text{Gray} \cdot \text{persons}}.$$

We simulated 100 different data sets. However, unperturbed data f_i, Q_i^{tr} and M_i^{mes}, $i = 1, \ldots, N$, are based on real data and are the same for all realizations. In fact,
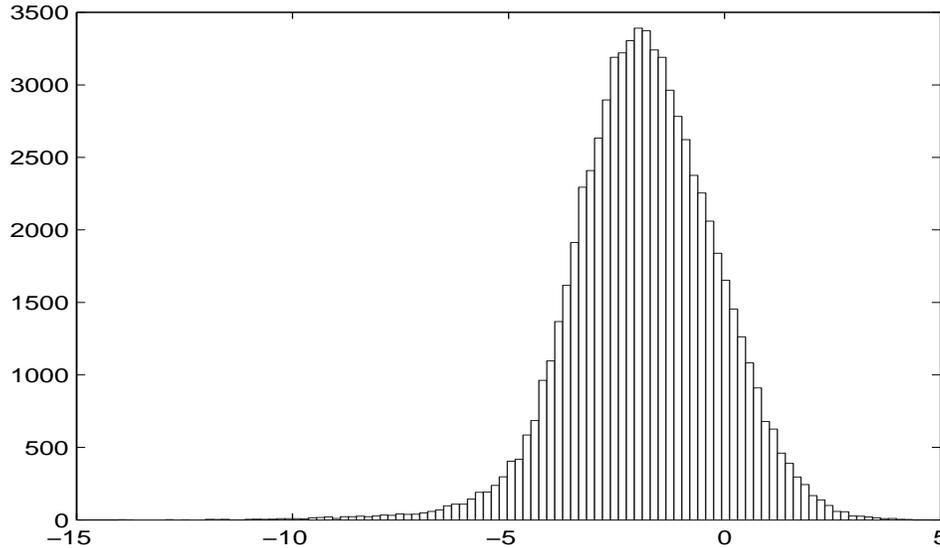
Figure 2: Uncensored distribution of $\log(fQ^{tr}/M^{mes})$.

- f_i is the estimate from the ecological dosimetric model, see Likhtarev et al. (2006b),
- for Q_i^{tr} we took the thyroid activity measured in 1986, see Likhtarev et al. (1993),
- and M_i^{mes} is the median of the thyroid mass distribution for corresponding age-sex-region group, see *Chornobyl: A Decade* (1997).

We used different values of the measurement error variances, including the real ones as obtained in *Chernobyl: A Decade* (1997), Likhtarev et al. (1993), and Likhtarev et al. (2003). The values of the geometric standard deviation GSD_M of the uncertainty distribution for the thyroid mass were 1 (no error), 1.33,[2] 1.5, and 2.0, while the GSD_Q of the uncertainty distribution for the measured activity varied from 1.5 to 5.0.

6.3   Simulation Results and Discussion

6.3.1   Setup and Methods

The simulation results are presented in Tables 1–4.

---

[2]Here the estimate of GSD_M based on the real data is used in the simulation of V_i^M. In this case the simulated error in the thyroid mass is heteroscedastic, and its average GSD is $\frac{1}{N}\sum_{i=1}^{N}\exp(\sigma_{M,i}) = 1.33$. For all other cases the simulated errors V_i^M are homoscedastic. Throughout the simulation, the errors V_i^Q are homoscedastic.

11

Table 1: Estimates of baseline incidence rate for the first scenario.

No.	GSD$_M$	GSD$_Q$	Naive (EPICURE)		Full ML				Regression Calibration			
					Parametric		Nonparametric		Parametric		Nonparametric	
			Mean	95% DI	Mean	95% CI	Mean	95% CI	Mean	95% DI	Mean	95% DI
1	No error	1.5	1.15	0.82-1.58	1.11	0.81-1.50	1.01	0.72-1.45	1.06	0.76-1.50	1.00	0.68-1.45
		2	1.40	1.03-1.85	1.28	0.95-1.68	1.02	0.71-1.45	1.17	0.84-1.62	1.00	0.64-1.47
		3	1.83	1.47-2.28	1.29	0.92-1.80	1.02	0.66-1.47	1.32	0.94-1.80	1.00	0.57-1.52
		5	2.22	1.83-2.71	1.33	0.89-1.84	1.03	0.53-1.62	1.42	1.00-1.93	1.05	0.50-1.83
2	Real esti-mates	1.5	1.15	0.79-1.58	1.11	0.79-1.52	1.01	0.64-1.41	1.07	0.76-1.50	1.00	0.68-1.46
		2	1.43	1.01-1.88	1.29	0.96-1.70	1.01	0.64-1.45	1.18	0.85-1.63	0.99	0.62-1.46
		3	1.89	1.51-2.36	1.29	0.95-1.77	1.01	0.68-1.47	1.33	0.98-1.82	1.00	0.51-1.52
		5	2.32	1.94-2.79	1.35	0.94-1.88	1.04	0.58-1.63	1.45	1.04-1.97	1.04	0.38-1.78
3	1.5	1.5	1.16	0.81-1.56	1.03	0.75-1.47	1.00	0.65-1.37	1.07	0.75-1.47	1.00	0.66-1.43
		2	1.45	1.08-1.90	1.30	0.91-1.76	1.02	0.70-1.46	1.19	0.88-1.62	0.99	0.67-1.45
		3	1.95	1.58-2.44	1.31	1.04-2.06	1.02	0.61-1.49	1.35	0.99-1.78	0.99	0.53-1.53
		5	2.42	2.03-3.00	1.37	0.95-1.74	1.03	0.55-1.60	1.48	1.03-1.96	1.03	0.39-1.78
4	2	1.5	1.20	0.82-1.70	1.03	0.68-1.52	1.00	0.67-1.46	1.08	0.75-1.56	0.99	0.63-1.47
		2	1.58	1.21-2.02	1.21	0.85-1.76	1.00	0.62-1.47	1.23	0.85-1.69	0.99	0.56-1.47
		3	2.26	1.87-2.71	1.40	1.01-1.85	1.00	0.52-1.52	1.44	1.01-1.90	0.99	0.47-1.52
		5	2.92	2.50-3.44	1.49	1.01-2.03	1.03	0.40-1.63	1.62	1.18-2.17	1.04	0.35-1.87

The header notes: **ERROR** | **Estimates of $\lambda_0 \times 10^3$ by different methods. True value of $\lambda_0 \times 10^3 = 1.0$**

Table 2: Estimates of excess absolute risk for the first scenario.

No.	ERROR		Estimates of EAR $\times 10^3$ by different methods. True value of EAR $\times 10^3 = 5.0$								
	GSD_M	GSD_Q	Naive (EPICURE)		Full ML				Regression Calibration			
					Parametric		Nonparametric		Parametric		Nonparametric	
			Mean	95% DI	Mean	95% CI	Mean	95% CI	Mean	95% DI	Mean	95% DI
1	No error	1.5	4.19	2.83-5.46	4.57	3.30-5.92	5.11	3.80-6.49	4.61	3.52-5.92	5.04	3.47-6.60
		2	2.91	1.89-3.95	3.88	2.92-5.28	5.11	3.57-6.49	4.00	2.88-5.21	5.04	3.44-6.63
		3	1.24	0.72-1.82	3.62	2.32-4.99	5.08	3.63-6.49	3.31	2.16-4.38	5.01	3.28-6.77
		5	0.26	0.07-0.47	3.23	1.94-4.50	5.04	3.78-6.70	2.85	1.83-4.02	4.81	2.65-6.91
2	Real esti-mates	1.5	4.60	3.14-6.22	3.23	1.94-4.50	5.08	3.53-6.11	4.62	3.47-6.21	5.05	3.53-6.75
		2	3.21	2.11-4.55	3.89	2.73-5.47	5.08	3.63-6.39	4.02	3.02-5.59	5.06	3.54-6.80
		3	1.38	0.81-2.13	3.63	2.66-5.06	5.05	3.66-6.51	3.33	2.32-4.72	5.03	3.55-6.73
		5	0.30	0.10-0.54	3.23	2.18-4.52	4.98	3.73-6.29	2.88	1.85-4.10	4.84	3.12-7.00
3	1.5	1.5	4.98	3.41-6.52	4.80	3.48-6.10	5.05	3.50-6.53	4.62	3.46-5.99	5.05	3.52-6.52
		2	3.49	2.28-4.68	3.95	2.62-5.35	5.01	3.50-6.35	4.03	3.05-5.41	5.05	3.52-6.61
		3	1.51	0.94-2.20	3.62	2.53-4.99	5.00	3.52-6.35	3.35	2.38-4.57	5.03	3.50-6.52
		5	0.33	0.11-0.55	3.23	1.01-4.36	4.94	3.59-6.35	2.90	1.97-3.98	4.86	3.28-6.73
4	2	1.5	6.88	5.22-8.44	4.71	3.61-5.78	5.00	3.85-6.14	4.64	3.53-5.63	5.04	3.87-6.09
		2	4.87	3.47-6.26	4.07	3.06-5.23	5.02	3.71-6.17	4.07	2.99-5.05	5.04	3.69-6.22
		3	2.14	1.39-2.94	3.65	2.65-4.69	5.01	3.54-6.43	3.40	2.38-4.37	5.02	3.68-6.46
		5	0.47	0.22-0.80	3.26	2.28-4.21	4.97	3.44-6.64	2.94	1.95-3.99	4.87	3.30-6.75

13

Table 1 (*continued*).

No.	GSD$_M$	GSD$_Q$	Estimates of $\lambda_0 \times 10^3$ by SIMEX			
	ERROR		Ordinary		Efficient	
			Mean	95% CI	Mean	95% CI
1	No error	1.5	1.00	0.60-1.40	1.00	0.62-1.39
		2	0.99	0.56-1.43	1.00	0.61-1.40
		3	1.11	0.62-1.61	1.02	0.60-1.44
		5	1.59	1.08-2.09	1.07	0.64-1.50
2	Real esti-mates	1.5	1.00	0.59-1.40	1.00	0.61-1.39
		2	0.98	0.54-1.42	0.99	0.59-1.40
		3	1.10	0.59-1.61	1.02	0.60-1.43
		5	1.60	1.09-2.11	1.07	0.63-1.51
3	1.5	1.5	0.99	0.59-1.40	1.00	0.60-1.39
		2	0.98	0.53-1.42	0.99	0.59-1.39
		3	1.09	0.57-1.62	1.01	0.58-1.44
		5	1.65	1.12-2.18	1.07	0.62-1.53
4	2	1.5	0.99	0.56-1.42	0.99	0.58-1.40
		2	0.95	0.48-1.42	0.99	0.56-1.41
		3	1.08	0.50-1.66	1.02	0.57-1.47
		5	1.83	1.23-2.44	1.13	0.65-1.61

Table 2 (*continued*).

No.	GSD$_M$	GSD$_Q$	Estimates of EAR $\times 10^3$ by SIMEX			
	ERROR		Ordinary		Efficient	
			Mean	95% CI	Mean	95% CI
1	No error	1.5	5.04	3.60-6.49	5.04	3.62-6.46
		2	4.94	3.44-6.43	5.04	3.60-6.49
		3	3.73	2.53-4.94	4.99	3.46-6.52
		5	1.39	0.86-1.93	4.83	3.25-6.41
2	Real esti-mates	1.5	5.06	3.68-6.43	5.06	3.71-6.40
		2	4.97	3.55-6.40	5.07	3.69-6.44
		3	3.80	2.64-4.96	5.01	3.58-6.43
		5	1.46	0.95-1.97	4.86	3.35-6.36
3	1.5	1.5	5.05	3.74-6.36	5.05	3.77-6.33
		2	4.97	3.63-6.32	5.06	3.76-6.37
		3	3.83	2.71-4.92	5.01	3.64-6.39
		5	1.45	0.96-1.95	4.85	3.40-6.30
4	2	1.5	5.05	3.97-6.13	5.05	4.00-6.10
		2	5.00	3.89-6.11	5.06	3.98-6.14
		3	3.92	2.97-4.87	4.99	3.87-6.11
		5	1.50	1.08-1.92	4.78	3.60-5.97

Table 3: Estimates of baseline incidence rate for the second scenario.

No.	GSD$_M$	GSD$_Q$	Naive (EPICURE)		Full ML				Regression Calibration			
					Parametric		Nonparametric		Parametric		Nonparametric	
			Mean	95% DI	Mean	95% CI	Mean	95% CI	Mean	95% DI	Mean	95% DI
1	No error	1.5	1.16	0.83-1.49	1.05	0.73-1.36	1.06	0.77-1.33	1.04	0.72-1.35	1.04	0.73-1.35
		2	1.39	1.06-1.73	1.07	0.75-1.38	1.04	0.71-1.34	1.05	0.73-1.36	1.04	0.75-1.34
		3	1.87	1.49-2.21	1.11	0.79-1.43	1.05	0.71-1.34	1.06	0.76-1.39	1.07	0.76-1.41
		5	2.56	2.10-2.96	1.16	0.80-1.50	1.03	0.70-1.39	1.13	0.75-1.49	1.06	0.74-1.46
2	Real esti-mates	1.5	1.21	0.85-1.54	1.10	0.73-1.42	1.06	0.76-1.41	1.04	0.66-1.41	1.09	0.73-1.42
		2	1.45	1.10-1.77	1.12	0.77-1.41	1.05	0.70-1.39	1.05	0.68-1.43	1.09	0.77-1.39
		3	1.97	1.59-2.29	1.16	0.82-1.44	1.04	0.75-1.35	1.07	0.71-1.43	1.10	0.81-1.35
		5	2.72	2.33-3.07	1.21	0.83-1.46	1.03	0.70-1.28	1.14	0.71-1.56	1.11	0.83-1.35
3	1.5	1.5	1.22	0.82-1.51	1.09	0.71-1.38	1.06	0.69-1.34	1.04	0.65-1.46	1.09	0.71-1.37
		2	1.48	1.08-1.79	1.11	0.75-1.42	1.04	0.69-1.33	1.05	0.67-1.47	1.10	0.69-1.40
		3	2.02	1.60-2.33	1.16	0.81-1.51	1.04	0.72-1.35	1.06	0.69-1.46	1.09	0.71-1.40
		5	2.81	2.43-3.14	1.21	0.87-1.52	1.01	0.63-1.33	1.13	0.70-1.52	1.09	0.75-1.46
4	2	1.5	1.25	0.80-1.60	1.10	0.66-1.43	1.06	0.64-1.37	1.05	0.66-1.43	1.09	0.66-1.42
		2	1.59	1.12-1.96	1.13	0.70-1.51	1.04	0.68-1.37	1.06	0.68-1.48	1.11	0.69-1.42
		3	2.32	1.87-2.71	1.19	0.75-1.55	1.03	0.65-1.49	1.08	0.73-1.49	1.12	0.75-1.43
		5	3.40	3.03-3.87	1.27	0.82-1.70	1.03	0.61-1.50	1.17	0.71-1.72	1.11	0.68-1.59

The table header: **Estimates of $\lambda_0 \times 10^3$ by different methods. True value of $\lambda_0 \times 10^3 = 1.0$**

Table 4: Estimates of excess absolute risk for the second scenario.

No.	ERROR		Estimates of EAR $\times 10^3$ by different methods. True value of EAR $\times 10^3 = 5.0$							
	GSD$_M$	GSD$_Q$	Naive (EPICURE)		Full ML				Regression Calibration			
					Parametric		Nonparametric		Parametric		Nonparametric	
			Mean	95% DI	Mean	95% CI	Mean	95% CI	Mean	95% DI	Mean	95% DI
1	No error	1.5	4.33	3.57-5.31	4.82	3.97-5.87	5.08	4.33-6.35	4.88	4.03-5.95	5.00	4.12-6.09
		2	3.39	2.81-4.21	4.69	3.88-5.72	5.17	4.40-6.44	4.84	4.04-5.96	4.97	4.09-6.10
		3	1.92	1.49-2.40	4.49	3.72-5.55	5.14	4.30-6.32	4.78	3.94-5.87	4.88	4.00-6.05
		5	0.62	0.44-0.82	4.30	3.56-5.34	5.12	4.16-6.63	4.57	3.73-5.75	4.81	3.91-5.93
2	Real esti-mates	1.5	4.71	3.92-5.63	4.77	3.96-5.68	5.01	4.18-5.97	4.86	4.11-5.74	4.93	4.07-5.85
		2	3.70	3.03-4.47	4.64	3.84-5.54	5.10	4.20-6.29	4.82	4.09-5.79	4.93	4.08-5.86
		3	2.10	1.71-2.57	4.46	3.75-5.31	5.11	4.12-6.17	4.76	3.98-5.74	4.86	4.17-5.79
		5	0.70	0.54-0.91	4.28	3.70-5.07	5.09	3.88-6.19	4.55	3.62-5.67	4.79	4.23-5.74
3	1.5	1.5	5.04	4.38-6.01	4.75	4.13-5.63	5.01	4.32-5.90	4.87	4.18-5.70	4.93	4.28-5.87
		2	3.96	3.40-4.75	4.63	3.97-5.48	5.09	4.32-6.17	4.80	4.09-5.73	4.90	4.25-5.81
		3	2.26	1.86-2.71	4.43	3.74-5.23	5.08	4.27-6.13	4.74	4.02-5.70	4.83	4.10-5.71
		5	0.76	0.60-0.98	4.25	3.66-4.98	5.08	4.24-6.38	4.55	3.76-5.57	4.76	4.12-5.64
4	2	1.5	6.87	5.92-7.75	4.70	4.06-5.28	4.95	4.24-5.50	4.80	4.11-5.34	4.87	4.17-5.44
		2	5.44	4.66-6.18	4.57	3.97-5.13	5.02	4.30-5.62	4.73	4.07-5.25	4.83	4.20-5.42
		3	3.12	2.62-3.62	4.37	3.86-4.91	5.02	4.23-5.62	4.66	4.00-5.22	4.74	4.25-5.26
		5	1.05	0.84-1.28	4.17	3.72-4.73	4.98	4.09-5.97	4.47	3.81-5.29	4.70	4.15-5.37

Table 3 (*continued*).

No.	GSD$_M$	GSD$_Q$	Estimates of $\lambda_0 \times 10^3$ by SIMEX			
			Ordinary		Efficient	
			Mean	95% CI	Mean	95% CI
1	No error	1.5	1.02	0.66-1.38	1.02	0.67-1.37
		2	1.01	0.64-1.39	1.02	0.66-1.37
		3	0.99	0.58-1.41	1.02	0.64-1.40
		5	1.15	0.65-1.64	1.05	0.65-1.46
2	Real esti-mates	1.5	1.02	0.66-1.38	1.02	0.66-1.38
		2	1.01	0.63-1.39	1.02	0.65-1.38
		3	0.98	0.55-1.42	1.02	0.64-1.40
		5	1.15	0.65-1.65	1.05	0.63-1.46
3	1.5	1.5	1.02	0.65-1.38	1.02	0.66-1.38
		2	1.01	0.62-1.39	1.01	0.65-1.38
		3	0.98	0.54-1.41	1.02	0.63-1.40
		5	1.15	0.64-1.66	1.06	0.64-1.48
4	2	1.5	1.02	0.63-1.41	1.03	0.64-1.41
		2	1.01	0.59-1.42	1.02	0.63-1.41
		3	0.94	0.47-1.41	1.02	0.61-1.43
		5	1.14	0.57-1.71	1.08	0.63-1.53

Table 4 (*continued*).

No.	GSD$_M$	GSD$_Q$	Estimates of EAR $\times 10^3$ by SIMEX			
			Ordinary		Efficient	
			Mean	95% CI	Mean	95% CI
1	No error	1.5	4.95	4.09-5.80	4.95	4.09-5.80
		2	4.91	4.04-5.77	4.95	4.08-5.82
		3	4.44	3.67-5.22	4.95	4.04-5.87
		5	2.65	2.20-3.11	4.90	3.88-5.92
2	Real esti-mates	1.5	4.93	4.11-5.74	4.93	4.12-5.74
		2	4.90	4.07-5.72	4.93	4.11-5.76
		3	4.44	3.69-5.19	4.94	4.07-5.81
		5	2.68	2.24-3.12	4.90	3.93-5.88
3	1.5	1.5	4.92	4.13-5.70	4.91	4.14-5.69
		2	4.88	4.09-5.67	4.92	4.13-5.72
		3	4.44	3.72-5.15	4.92	4.09-5.76
		5	2.68	2.26-3.09	4.87	3.93-5.81
4	2	1.5	4.85	4.19-5.51	4.84	4.19-5.50
		2	4.82	4.15-5.50	4.85	4.18-5.52
		3	4.42	3.81-5.03	4.85	4.15-5.56
		5	2.70	2.33-3.07	4.80	4.00-5.60

The estimation of the absolute risk parameters was performed by several methods:

1. Naïve: The ordinary maximum likelihood method in which the dose estimates are taken to be error-free.
2. Parametric Full ML: Full maximum likelihood where the error in the measured activity was taken into account by means of the integral convolution over the distribution of true thyroid doses (Carroll et al., 2006; Masiuk et al., 2008), assuming that fQ^{tr}/Mmes has a lognormal distribution. The error in the thyroid mass was handled by relation (13).
3. Nonparametric Full ML: The full maximum likelihood method described above, but without assumption about the distribution of fQ^{tr}/Mmes. The latter distribution is approximated according to formula (12).
4. Parametric Regression Calibration: Parametric regression calibration as described in Section 4.1.
5. Nonparametric Regression Calibration: Nonparametric calibration as described in Section 4.2.
6. Ordinary SIMEX: The SIMEX method described in Section 5 with refinement (A.3).
7. Efficient SIMEX: Our computationally efficient modification of the SIMEX method presented in Section 5.

Each estimate was computed for 100 different realizations of doses and cases. Then the corresponding estimates were averaged.

For methods 1, 4 and 5, the confidence intervals were not calculated due to their well-known statistical inconsistency in the presence of dose errors, and hence we used the deviance interval (DI), calculated based on the 2.5th and the 97.5th percentiles of the estimates over 100 simulations.

For the estimators 2, 3, 6 and 7 the confidence intervals were constructed based on the asymptotic covariance matrices. The method applied to construct the confidence intervals for both SIMEX estimators is given in Appendix A.3.

6.3.2 The Naïve Method

The analysis of the results shows that in cases where the errors are present only in the measured activities (see the first block in Tables 1-4), the naïve method underestimates of the excess absolute risk, while the baseline incidence rate is overestimated. The bias increases as the errors increase. For $GSD_Q = 5$ the value of the EAR is underestimated by more than a factor of 10.

If the errors are present both in the measured activity and in the thyroid mass, then the naïve estimate of the excess absolute risk approaches the true value only in case where the error variances are approximately equal. This is connected to the fact that Berkson errors in M_i^{tr} lead to the overestimation of the EAR, while the classical errors lead conversely to its underestimation. Thus, the errors compensate for one another. If one of the errors is larger, then the EAR estimator is biased.

6.3.3 Parametric Regression Calibration and Maximum Likelihood

In Scenario 1, parametric methods are biased (see Tables 1 and 2), because the distribution of the thyroid doses is not lognormal. In other words, the misspecification of the true distribution leads to bias of the estimates which are constructed by structural methods (Schneeweiss and Cheng, 2006).

In Scenario 2 (see Tables 3 and 4), the dose distribution is uncensored and can be approximated by the lognormal law. It thus comes as no surprise that the parametric methods therefore have little bias.

6.3.4 Nonparametric Regression Calibration and Maximum Likelihood

Nonparametric regression calibration and nonparametric full maximum likelihood methods seem to be more adequate, because they estimate the dose distribution in a nonparametric way. As a result those methods show quite modest bias even for the nonsmooth underlying distribution in Scenario 1.

However, in Scenario 2, things might be expected to be different because of the large right tail of true doses, for which the regression calibration approximation might break down. Nonparametric full maximum likelihood behaves quite well in this case, while nonparametric regression calibration has some bias, perhaps because the nonlinear effects of the underlying logistic type model are revealed.

6.3.5 SIMEX

Though SIMEX is robust with respect to the dose distribution, of course its performance gets worse for large errors. In particular ordinary SIMEX shows considerable bias for large measurement errors. The reason is as follows: we use a quadratic extrapolation function, but the naive ordinary likelihood estimator deviates from the quadratic law as a function of the additional error variance. In the computationally efficient SIMEX procedure, we use modified naive estimators given by the estimating equations (15) and (16), and those naive estimators vary almost as a quadratic function of the additional error variance. As a result the efficient SIMEX has small

19

bias even for relatively large errors. In the simulation the level of GSD_Q was limited to 5, but for larger classical errors the behavior of the computationally efficient SIMEX estimator could be worse.

6.3.6 Numerical Comparisons

For a particular comparison, we consider the case that $GSD_Q = 3$ and $GSD_M = 1.5$.

In Scenario 1, the %-bias in λ_0 for Naïve, Parametric ML, Nonparametric ML, Parametric regression calibration, Nonparametric regression calibration, Ordinary SIMEX and Efficient SIMEX were 95%, 31%, 2%, 35%, 1%, 9% and 1%, respectively. The corresponding biases for Scenario 2 were 102%, 16%, 4%, 6%, 9%, 2% and 2%, respectively.

In Scenario 1, the %-bias in Excess Absolute Risk for Naïve, Parametric ML, Nonparametric ML, Parametric regression calibration, Nonparametric regression calibration, Ordinary SIMEX and computationally Efficient SIMEX were 70%, 28%, 0.1%, 13%, 1%, 53% and 0.2%, respectively. The corresponding biases for Scenario 2 were 50%, 11%, 2%, 5%, 3%, 11% and 2%, respectively.

In terms of variability, we ignore the Naïve estimate because of its severe bias. We considered the length of the 95% CI as a measure of standard deviation, and define the variance efficiency of a method compared to the computationally efficient SIMEX method as the square of the ratio of efficient SIMEX interval length to the method's interval length. For the methods the CI were not computed, the deviance intervals were used.

In Scenario 1, the variance efficiency in λ_0 for the Parametric ML, Nonparametric ML, Parametric regression calibration, Nonparametric regression calibration and Ordinary SIMEX compared to computationally Efficient SIMEX were 100%, 71%, 59%, 118% and 73%, respectively. The corresponding efficiencies for Scenario 2 were 121%, 150%, 100%, 125% and 79%, respectively.

In Scenario 1, the variance efficiency in Excess Absolute Risk for the Parametric ML, Nonparametric ML, Parametric regression calibration, Nonparametric regression calibration and Ordinary SIMEX compared to computationally Efficient SIMEX were 125%, 94%, 158%, 83% and 155%, respectively. The corresponding efficiencies for Scenario 2 were 125%, 81%, 99%, 107% and 137%, respectively.

6.3.7 Other Comments

Simulations showed that the level of the classical and Berkson errors has little influence on the size of confidence intervals for the computationally efficient SIMEX

and nonparametric full maximum likelihood estimates. Nevertheless for larger classic errors confidence intervals become wider. Moreover, in some cases larger Berkson errors correspond to narrower confidence intervals. The reason for this effect is that in the presence of a larger Berkson error, the number of disease cases in the simulated response increases because the mean dose is increasing. This in turn yields more information and leads to shrinking of confidence intervals for risk estimators.

We considered model (1) in terms of the baseline incidence rate and EAR because we think it is more natural in the measurement error context. Additionally, EAR is a kind of slope parameter for dose, and the attenuation effect for the naïve estimator of slope parameter is well known (Carroll et al., 2006). But it is possible to compute point estimates of ERR as well and construct the corresponding confidence intervals. A rough estimate of ERR equals the ratio of estimates of EAR and λ_0.

7 Conclusions

We studied the influence of mixed dose errors on radiation risk estimation in binary regression with linear dose dependence. A model of absolute risk was investigated where the doses are observed with a mixture of classical and Berkson multiplicative errors. The algorithms were constructed to compute the parameter estimates by full maximum likelihood, regression calibration and SIMEX method.

For the standard binary model, i.e., under the absence of errors in the dose estimates, a computationally efficient algorithm to compute the risk estimate was proposed, which leads to solving a one variable equation. In Appendix A.1 it is shown that this equation has no more than one root in the region $\lambda_0 > 0$, $\beta > 0$ and it can be solved efficiently. Simulations show that it is almost as accurate as the maximum likelihood estimator and can serve as a good initial approximation for computation of the maximum likelihood estimator. Our simulations suggest that for the range of doses considered, the efficient SIMEX estimators turned out to be the most precise of all the estimators we considered.

We studied two scenarios in our simulation, with the following ideas in mind. Parametric regression calibration and parametric maximum likelihood necessarily make parametric assumptions about the distribution of a latent variable fQ^{tr}/Mmes. We were interested in what happens when the parametric assumptions were very wrong (Scenario 1, Tables 1 and 2) or were approximately correct (Scenario 2, Tables 3 and 4). Not surprisingly, when the assumptions were very wrong, the parametric methods incurred significant bias. This shows that parametric regression calibration and maximum likelihood are excellent procedures as long as

21

care is taken in the modeling assumptions. Whether Scenario 1 or Scenario 2 is the more practical case will depend on the application.

Among the methods that make no assumptions about the distribution of the latent variable, we generally observed little bias. Although the computationally efficient SIMEX and nonparametric full maximum likelihood seem to be more accurate, the nonparametric regression calibration is easier to apply. Indeed within the nonparametric calibration procedure, after estimating the underlying dose distribution and calibrating the doses, the standard package *EPICURE* produces the final risk estimates.

Appendix: Sketch of Technical Arguments

A.1   Consistency of the Computationally Efficient Methods

Introduce a function

$$F_N(\beta) = \left(\Sigma_{i=1}^{N} \frac{Y_i}{1+\beta D_i} \right)^{-1} \Sigma_{i=1}^{N} \frac{Y_i(D_i - D_{\mathrm{av}})}{1+\beta D_i}, \quad \beta \geq 0,$$

where D_{av} is defined by equation (5). The relationship (4) is equivalent to the equation $F_N(\beta) = 0, \quad \beta > 0$. Assume the following:

(i)   Among the observations with $Y_i = 1$, not all the values D_i coincide.
(ii) $\Sigma_{i=1}^{N} Y_i(D_i - D_{\mathrm{av}}) > 0$.
(iii) $\Sigma_{i=1}^{N} Y_i(D_i - D_{\mathrm{av}})/D_i < 0$.

Condition (i) means that not all ill subjects received an identical dose. Condition (ii) is natural as well and means that the mean dose for healthy subjects is less than the mean dose for ill subjects. It is more difficult to interpret condition (iii). One can show that for the structural model where the dose distribution does not degenerate to a single point (i.e. for each $D_0 \geq 0$, it holds $\mathrm{pr}(D = D_0) < 1$), the condition (iii) holds true with probability 1 for all N starting from a certain random number.

<u>Lemma 1</u> *1. Assume condition (i). Then the function F_N is strictly decreasing, and therefore, equation (4) has no more than one solution.*

2. Assume conditions (i) to (iii). Then there exists a unique solution $\beta > 0$ to equation (4).

The proof of 1. consists in verifying inequality $F_N'(\beta) < 0$, $\beta \geq 0$. Next, (ii) implies that $F_N(0) > 0$, and from (iii) it follows that $\lim_{\beta \to +\infty} F_N(\beta) < 0$, and then a continuous function F_N equals zero at certain point $\beta > 0$.

Theorem 1 *Assume that in the structural model (1) the distribution of dose D is not concentrated at a single point. Then the computationally efficient estimates* $\widehat{\lambda}_{0Q}$, $\widehat{\beta}_Q$ *exist and are unique for all* $N \geq N_0(\omega)$ *almost surely, moreover those estimates are strictly consistent, i.e. with probability 1 they converge to the true values* $\lim_{N \to \infty} \widehat{\lambda}_{0Q} = \lambda_0$ *and* $\lim_{N \to \infty} \widehat{\beta}_Q = \beta$, *almost surely.*

Here is a sketch proof of Theorem 1. Lemma 1 implies the existence of the estimates for $N \geq N_0(\omega)$, almost surely. The estimating equations for the efficient estimator can be rewritten as

$$
\begin{aligned}
0 &= \sum_{i=1}^{N} [1 - Y_i - Y_i/\{\lambda_0(1 + \beta D_i)\}], & \text{(A.1)} \\
0 &= \sum_{i=1}^{N} D_i [1 - Y_i - Y_i/\{\lambda_0(1 + \beta D_i)\}] & \text{(A.2)}
\end{aligned}
$$

(over the domain $\{\lambda_0 > 0, \ \beta > 0\}$ these equations are equivalent to (2)–(3); and to obtain equation (A.2), one has to subtract (3) from (2) and divide the difference by β). Equations (A.1)–(A.2) are unbiased, that is

$$
\begin{aligned}
0 &= \mathbf{E}_{\lambda_0,\beta} [1 - Y - Y/\{\lambda_0(1 + \beta D)\}]; \\
0 &= \mathbf{E}_{\lambda_0,\beta} (D[1 - Y - Y/\{\lambda_0(1 + \beta D)\}]).
\end{aligned}
$$

Here and hereafter $\mathbf{E}_{\lambda_0,\beta}$ denotes the expectation, provided λ_0 and β are true values of the parameters in the model (1). Moreover, the limit system of equations

$$
\begin{aligned}
0 &= \mathbf{E}_{\lambda_0,\beta} [1 - Y - Y/\{\ell_0(1 + bD)\}]; \\
0 &= \mathbf{E}_{\lambda_0,\beta} (D[1 - Y - Y/\{\ell_0(1 + bD)\}]); \quad \ell_0, b > 0
\end{aligned}
$$

has the unique solution $\ell_0 = \lambda_0$, $b = \beta$. This can be proven using the convex function $q(\ell_0, a) = (1 - Y)(\ell_0 + aD) - Y \ln(\ell_0 + aD)$. The functions under expectation are its partial derivatives at point $(\ell_0, a) = (\ell_0, \ell_0 b)$. By the theory of estimating equations (Carroll et al, 2006; Huber, 1967) the last two facts imply the strict consistency of the efficient estimates.

A.2   Computationally Efficient SIMEX

A.2.1   Description of the Estimator

The SIMEX estimator described in Section 5 differs from ordinary SIMEX as follows.

23

1. A model of observations of activities Q_i with heteroscedastic errors is considered, i.e., the variances of the classical errors $\sigma^2_{Q,i}$ can vary with the trial number i. At that the variances are assumed known and are not the values of, e.g., a known function of the unobserved value Q_i^{tr}.
2. The estimator given by equations (2), (3) plays the role of the naïve maximum likelihood estimator for different values of κ. This simplifies calculations without essential loss in accuracy of estimation.
3. Perturbations $U^*_{b,i}$ are simulated in such a way that

$$\sum_{b=1}^{B} U^*_{b,i} = 0, \quad i = 1,...,N. \tag{A.3}$$

We now explain why equality (A.3) provides smaller variability and smaller deviations.

Let $\theta = (\lambda_0, EAR)^{\mathrm{T}}$, $s(\theta, Y, D)$ be the elementary estimating function for the naïve estimate, $\widehat{\theta}^*_b(\kappa)$ and $\widehat{\theta}^*(\kappa)$ the estimates of the parameter θ used in the SIMEX method. For ordinary SIMEX estimator which is combined with efficient estimates, the elementary estimation function can be written in the form, see equations (A.1)–(A.2),

$$s(\theta, Y, D) = (1, D)^{\mathrm{T}} \left\{ 1 - Y - Y/(\lambda_0 + EAR \cdot D) \right\},$$

while for the efficient modification of SIMEX

$$s(\theta, Y, D) = \left\{ 1 - Y - Y/(\lambda_0 + EAR \cdot D), \ (1 - Y)\widehat{D}_{\mathrm{av}} - YD/(\lambda_0 + EAR \cdot D) \right\}^{\mathrm{T}},$$

where $\widehat{D}_{\mathrm{av}} = \{\sum_{i=1}^{N}(1 - Y_i)\}^{-1}\sum_{i=1}^{N}(1 - Y_i)D_i^{Mcal}e^{-\frac{1}{2}\sigma^2_{Q,i}}$,
and $D_i^{Mcal} = (Q_i^{mes}/M_i^{mes})f_i e^{\frac{1}{2}\sigma^2_{M,i}}$.

The random function $\widehat{\theta}^*_b(\kappa)$ is found from the equation

$$\sum_{i=1}^{N} s(\widehat{\theta}^*_b(\kappa), Y_i, D_i^{Mcal}\exp(\sqrt{\kappa}U^*_{b,i})) = 0.$$

We find the derivative $\left.\frac{d\widehat{\theta}^*_b(r^2)}{dr}\right|_{r=0}$ with $r = \sqrt{\kappa}$. For this purpose we calculate the partial derivatives s'_D and s'_θ of the estimating function $s(\theta, Y, D)$:

$$\left.\frac{\partial \sum_{i=1}^{N} s(\theta, Y_i, D_i^{Mcal}e^{rU^*_{b,i}})}{\partial r}\right|_{r=0} = \sum_{i=1}^{N} s'_D(\theta, Y_i, D_i^{Mcal})D_i^{Mcal}U^*_{b,i};$$

$$\frac{\partial \sum_{i=1}^{N} s(\theta, Y_i, D_i^{Mcal} e^{rU_{b,i}^*})}{\partial \theta}\bigg|_{r=0} = \sum_{i=1}^{N} s_\theta'(\theta, Y_i, D_i^{Mcal})$$

$$= \sum_{i=1}^{N} \{Y_i/(\theta_1 + \theta_2 D_i^{Mcal})^2\} \begin{pmatrix} 1 & D_i^{Mcal} \\ D_i^{Mcal} & (D_i^{Mcal})^2 \end{pmatrix}.$$

Under the condition of existence of the computationally efficient estimate (see Appendix A.1) the derivative $\sum_{i=1}^{N} s_\theta'(\theta, Y_i, D_i^{Mcal})$ is a nonsingular matrix.

Then by the Implicit Function Theorem

$$\frac{d\widehat{\theta}_b^*(r^2)}{dr}\bigg|_{r=0} = -\left\{\sum_{i=1}^{N} s_\theta'(\widehat{\theta}_b^*(0), Y_i, D_i^{Mcal})\right\}^{-1} \sum_{i=1}^{N} s_D'(\widehat{\theta}_b^*(0), Y_i, D_i^{Mcal})D_i^{Mcal}U_{b,i}^*.$$

Note that $\widehat{\theta}_b^*(0) = \widehat{\theta}_1^*(0)$ does not depend on b and find the derivative of the function $\widehat{\theta}^*(r^2) = B^{-1}\sum_{b=1}^{B} \widehat{\theta}_b^*(r^2)$ at point 0:

$$\frac{d\widehat{\theta}^*(r^2)}{dr}\bigg|_{r=0} = -\frac{\sum_{i=1}^{N} s_D'(\widehat{\theta}_1^*(0), Y_i, D_i^{Mcal})D_i^{Mcal}\sum_{b=1}^{B} U_{b,i}^*}{B\sum_{i=1}^{N} s_\theta'(\widehat{\theta}_1^*(0), Y_i, D_i^{Mcal})}$$

If $\sum_{b=1}^{B} U_{b,i}^* = 0$ then $\{d\widehat{\theta}^*(r^2)/dr\}_{r=0} = 0$, and in the expansion

$$\widehat{\theta}^*(\kappa) = \widehat{\theta}^*(0) + coef_1\sqrt{\kappa} + coef_2\kappa + coef_3\sqrt{\kappa^3} + ... \qquad (A.4)$$

the term $coef_1\sqrt{\kappa}$ is vanishing.

Condition (A.3) decreases variability of the estimator. Indeed, under (A.3) the term $coef_1\sqrt{\kappa}$ in the expansion (A.4) is vanishing, and this expansion looks more like Taylor expansion w.r.t. κ. Therefore, the extrapolated value $\widehat{\theta}^*(-1)$, which is the SIMEX estimator, will be more numerically stable.

A.2.2 Computationally Efficient SIMEX as a Combination of the SIMEX Method and the Corrected Score Method

The expression $\sum_{i=1}^{N}(1-Y_i)Q_i^{mes}f_i\exp\{(\sigma_{M,i}^2 - \sigma_{Q,i}^2)/2\}/M_i^{mes} = \widehat{D}_{av}\sum_{i=1}^{N}(1-Y_i)$ is an unbiased estimator of $\sum_{i=1}^{N}(1-Y_i)D_i^{tr}$, where D_i^{tr} is defined in (9). Indeed,

$$\mathbf{E}[(1-Y_i)Q_i^{mes}f_i\exp\{(\sigma_{M,i}^2 - \sigma_{Q,i}^2)/2\}/M_i^{mes} \mid Y_i, Q_i^{tr}, M_i^{mes}, f_i]$$
$$= (1-Y_i)f_i\mathbf{E}[Q_i^{mes}\exp(-\sigma_{Q,i}^2/2) \mid Q_i^{tr}]\exp(\sigma_{M,i}^2/2)/M_i^{mes}$$
$$= (1-Y_i)f_iQ_i^{tr}\mathbf{E}[V_i^Q\exp(-\sigma_{Q,i}^2/2)]\mathbf{E}[(V_i^M)^{-1}]/M_i^{mes}$$
$$= (1-Y_i)f_iQ_i^{tr}\mathbf{E}[(M_i^{tr})^{-1} \mid M_i^{mes}]$$
$$= \mathbf{E}[(1-Y_i)D_i^{tr} \mid Y_i, Q_i^{tr}, M_i^{mes}, f_i].$$

We derived this equality using the model of observations (7)-(8) and the assumption on the lognormality of multiplicative errors.

Therefore, equation (15), which can be rewritten as

$$\sum_{i=1}^{N}(1-Y_i)(1+\beta\widehat{D}_{\text{av}}) = \sum_{i=1}^{N}Y_i/\lambda_0,$$

where the unknown variables are denoted as λ_0 and β while in (15) the same variables are denoted as $\widehat{\lambda}_{0,b}^{*}(\kappa)$ and $\widehat{\beta}_{b}^{*}(\kappa)$, is constructed from

$$\sum_{i=1}^{N}(1-Y_i)(1+\beta D_i^{tr}) = \sum_{i=1}^{N}Y_i/\lambda_0,$$

see equation (2), according to the Corrected Score method (Carroll et al., 2006). Like the SIMEX method, this method is functional and does not need the knowledge of the distribution of the latent variable.

A.3 Implementation of the Computationally Efficient SIMEX Method

Perturbations $U_{b,i}^{*} = \text{Normal}(0, \sigma_{Q,i}^2)$, $b = 1, \dots, B$, $B = 10$, were simulated in such a way that for the fixed i, the random variables $U_{b,i}^{*}$ have joint Gaussian distribution with the covariance

$$\mathbf{E}(U_{b_1,i}^{*}U_{b_2,i}^{*}) = -\sigma_{Q,i}^2/(B-1), \quad b_1 \neq b_2,$$

and therefore, $\sum_{b=1}^{B}U_{b,i}^{*} = 0$, $i = 1, ..., N$. The numerical extrapolation of the functions $\widehat{\lambda}_0^{*}(\kappa)$ and $EAR^{*}(\kappa)$ was performed via approximation by the least-squares polynomial of the second order at points $\kappa = 0$; 0.2; 0.4; 0.6. The explicit formula for the value of the extrapolating polynomial at point $\kappa = -1$ is as follows:

$$\widehat{\theta}^{*}(-1) = 12.45\,\widehat{\theta}^{*}(0) - 9.35\,\widehat{\theta}^{*}(0.2) - 10.65\,\widehat{\theta}^{*}(0.4) + 8.55\,\widehat{\theta}^{*}(0.6).$$

Here $\widehat{\theta}^{*}(\kappa)$ stands for either $\widehat{\lambda}_0^{*}(\kappa)$ or $EAR^{*}(\kappa)$.

The confidence interval for the SIMEX estimator is constructed based on the SIMEX Sandwich covariance matrix of this estimator. The covariance matrix is computed according to the recommendation in Carroll et al. (2006, pp. 111 and 393–395). We estimate the variance of the naive estimator $\widehat{\theta}_b^{*}(\kappa)$ taking Y_i and $D_{b,i}^{*}(\kappa)$ as the data using sandwich formula, and denote the variance estimate as $\widehat{\tau}_b^2(\kappa)$. Then SIMEX Sandwich variance estimator is the extrapolated value of $\frac{1}{B}\sum_{b=1}^{B}\widehat{\tau}_b^2(\kappa) - \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\theta}_b^{*}(\kappa) - \widehat{\theta}^{*}(\kappa))^2$ to the point $\kappa = -1$.

References

Anderson T. W. (1958): *An introduction to multivariate statistical analysis*, New York: John Wiley and Sons.

Apanasovich T. V., Carroll R. J., and Maity A. (2009): "SIMEX and variance estimation in semiparametric measurement error models," *Electronic Journal of Statistics*, **3**, 318–348.

Carroll R. J., Ruppert D., Stefanski L. A., and Crainiceanu C. A. (2006): *Measurement Error in Nonlinear Models. A Modern Perspective*, Boca Raton: Chapman and Hall/CRC.

Chernobyl: A Decade (1997). Proceedings of the Fifth Chernobyl Sasakawa Medical Cooperation Symposium, Kiev, Ukraine, 14–15 October 1996 (Eds.: Yamashita S., Shibata Y.) Amsterdam: Elsevier.

Health Risks from Exposure to Low Levels of Ionizing Radiation: (BEIR VII Phase 2, 2006), Washington, D.C.: National Academy Press.

Hofer E. (2008): "How to account for uncertainty due to measurement errors in an uncertainty analysis using Monte Carlo simulation," *Health Physics*, **95**, 277–290.

Hosmer D. and Lemeshow S. (2000): *Applied Logistic Regression*, New-York: John Wiley and Sons.

Huber P. J. (1967): "The behavior of maximum likelihood estimates under nonstandard conditions," *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, Berkeley: University of California Press, Vol. **1**, 221–233.

Jacob P., Bogdanova T. I., Buglova E., Chepurniy M., Demidchik Y., Gavrilin Y., Kenigsberg J., Meckbach R., Schotola C., Shinkarev S., Tronko M. D., Ulanovsky A., Vavilov S., and Walsh L. (2006): "Thyroid cancer risk in areas of Ukraine and Belarus affected by the Chernobyl accident," *Radiation Research*, **165**, 1–8.

Kopecky K. J., Stepanenko V., Rivkind N., Voilleque P., Onstad L., Shakhtarin V., Parshkov E., Kulikov S., Lushnikov E., Abrosimov A., Troshin V., Romanova G., Doroschenko V., Proshin A., Tsyb A., and Davis S. (2006): "Childhood thyroid cancer, radiation dose from Chernobyl and dose uncertainties in Bryansk Oblast, Russia: A population-based case-control study," *Radiation Research*, **166**, 367–374.

Kukush A. and Schneeweiss H. (2005): "Comparing different estimators in a nonlinear measurement error model. I, II," *Mathematical Methods of Statistics*, **14**, No 1, 53–79 and **14**, No 2, 203–223.

Kukush A., Malenko A., and Schneeweiss H. (2009): "Optimality of the quasi score estimator in a mean-variance model with applications to measurement error models," *Journal of Statistical Planning and Inference*, **139**, 3461–3472.

Likhtarev I. A., Prohl G. and Henrichs K. (1993): "Reliability and accuracy of the ^{131}I thyroid activity measurements performed in the Ukraine after the Chernobyl accident in 1986," Munich: GSF-Bericht 19/93, Institut für Strahlenschutz.

Likhtarev I., Minenko V., Khrouch V., and Bouville A. (2003): "Uncertainties in thyroid dose reconstruction after Chernobyl," *Radiation Protection Dosimetry*, **105**(1-4), 601–608.

Likhtarev I., Kovgan L., Vavilov S., Chepurny M., Bouville A., Luckyanov N., Jacob P., Voillequé P., and Voigt G. (2005): "Post-Chornobyl thyroid cancers in Ukraine. Report 1: Estimation of thyroid doses," *Radiation Research*, **163**, 125–136.

Likhtarev I., Kovgan L., Vavilov S., Chepurny M., Bouville A., Luckyanov N., Jacob P., Voillequé P., and Voigt G. (2006a) "Post-Chornobyl thyroid cancers in Ukraine. Report 2: Risk analysis," *Radiation Research*, **166**, 375–386.

Likhtarev I., Bouville A., Kovgan L., Luckyanov N., Voilleque P., and Chepurny M. (2006b): "Questionnaire- and Measurement-Based Individual Thyroid Doses in Ukraine Resulting from the Chornobyl Nuclear Reactor Accident," *Radiation Research*, **166**, 271–286.

Li Y., Guolo A., Hoffman F. O., and Carroll R. J. (2007): "Shared uncertainty in measurement error problems, with application to Nevada Test Site fallout data," *Biometrics*, **63**, 1226–1236.

Lyon J. L., Alder S. C., Stone M. B., Scholl A., Reading J. C., Holubkov R., Sheng X., White G. L., Hegmann K. T., Anspaugh L., Hoffman F. O., Simon S. L., Thomas B., Carroll R. J., and Meikle A. W. (2006): "Thyroid disease associated with exposure to the Nevada Test Site radiation: a reevaluation based on corrected dosimetry and examination data," *Epidemiology*, **17**, 604–614.

Mallick B., Hoffman F. O. and Carroll R. J. (2002): "Semiparametric regression modeling with mixtures of Berkson and classical error, with application to fallout from the Nevada Test Site," *Biometrics*, **58**, 13–20.

Masiuk S. V., Shklyar S. V., Kukush A. G., and Vavilov S. E. (2008): "Impact of doses uncertainties on the radiation risks estimation," *Radiation and Risk*, Bulletin of the National Radiation and Epidemiological Registry, **17**, No 3, 64–75 (in Russian). `http://mechmat.univ.kiev.ua/eng/ppages/kukush/Public/files/Unsert_risk_influence.doc`

Preston D. L., Lubin J. H., Pierce D. A. and McConney M. E. (1993): *EPICURE User's Guide*, Seattle, WA: Hirosoft Corporation.

Reeves G. K., Cox D. R., Darby S. C., and Whitley E. (1998): "Some aspects of measurement error in explanatory variables for continuous and binary regression models," *Statistics in Medicine*, **17**, 2157–2177.

Ron E. and Hoffman F. O. (1999): "Uncertainties in radiation dosimetry and their impact on dose-response analysis": Proc. of a workshop held September 3-5, 1997 in Bethesda, Maryland. – NIH Publication No. 99-4541.

Schneeweiss H. and Cheng C.-L. (2006): "Bias of the structural quasi-score estimator of a measurement error model under misspecification of the regressor distribution," *Journal of Multivariate Analysis*, **97**, 455–473.

Schneeweiss H. and Kukush A. (2009): "Comparing the efficiency of structural and functional methods in measurement error models," *Probability Theory and Mathematical Statistics* (*Teor. Ǐmovir. Matem. Statist.*, ISSN 0868-6904), **80**, 119–129. Reprinted in *Theory of Probability and Mathematical Statistics* (ISSN 1547-7363, ISSN 0094-9000), **80**, 131–142.

Shklyar S. and Schneeweiss H. (2005): "A comparison of asymptotic covariance matrices of three consistent estimators in the Poisson regression model with measurement errors," *Journal of Multivariate Analysis*, **94**, 250–270.

Stram D. O. and Kopecky K. J. (2003): "Power and uncertainty analysis of epidemiological studies of radiation-related disease risk in which dose estimates are based on a complex dosimetry system: some observations," *Radiation Research*, **160**, 408–417.

Tronko M. D., Howe G. R., Bogdanova T. I., Bouville A. C., Epstien O. V., Brill A. B., Likhtarev I. A., Fink D. J., Markov V. V., Greenebaum E., Olijnyk V. A., Masnyk I. J., Shpak V. M., McConnell R. J., Tereshchenko V. P., Robbins J., Zvinchuk O. V., Zablotska L. B., Hatch M., Luckyanov N. K., Ron E., Thomas T. L., Voillequé P. G., and Beebe G. W. (2006): "A cohort study of thyroid cancer and other thyroid diseases after the Chornobyl accident: thyroid cancer in Ukraine detected during first screening," *Journal of the National Cancer Institute*, **98**, 897–903.