RECOMMENDATIONS ON STATISTICAL APPROACHES TO ACCOUNT FOR DOSE UNCERTAINTIES IN RADIATION EPIDEMIOLOGIC RISK MODELS

Draft: September 28, 2023
Preface

Accurate exposure estimation in radiation epidemiologic studies is essential for reliable health risk assessment. Failure to account appropriately for dose uncertainties and model assumptions could lead to biased results in dose-response analyses and interpretation as well as incorrect confidence bounds for risk parameter estimates. Assessment of absorbed dose is often subject to considerable uncertainties, and a variety of statistical approaches have been developed to incorporate dose uncertainties into the dose-response models. The purpose of this Commentary is to provide a guide to available statistical methods for dose-response analysis that incorporate dose uncertainties, the types of studies to which the methods can be applied, and the advantages and drawbacks of the methods. This Commentary addresses studies of external and internal exposures and provides guidance on both shared and unshared uncertainty in the calculation of absorbed dose and the associated statistical uncertainties. Of particular interest are statistical methods of assessing dose-response in epidemiologic studies of internal emitters for which doses are calculated using exposure and retention models with many parameters, and each parameter has uncertainties.

This Commentary draws from and builds upon previous NCRP reports on closely related topics, including:


The intended audience for this Commentary is dosimetrists, statisticians, epidemiologists and organizations concerned with evaluating health risks associated with exposure of cohorts to radiation.

This Commentary was prepared by Scientific Committee 1-28 on Recommendations on Statistical Approaches to Account for Dose Uncertainties in Radiation Epidemiologic Risk Models. Serving on Scientific Committee 1-28 were:
The Council wishes to express its appreciation to the Committee members for the time and effort devoted to the preparation of this Commentary. NCRP could not continue to address the radiation protection needs of the nation without the willingness of Council members and others to serve, review and advise and without the partnership of agencies to work together for the good of the people and the well-being of the nation.

NCRP gratefully acknowledges the financial support provided for this effort by the U.S. Department of Energy under Grant No. DE-FG01-06-EH-06013.
## Contents

### Preface

1. Executive Summary
   1.1 Background 6

2. Introduction 10

3. Definitions of Uncertainty and Error Types
   3.1 Classical or Berkson Error 13
   3.2 Shared and Unshared Error 15
   3.3 Other Types of Uncertainty 16

4. Dosimetric Uncertainties in Radioepidemiologic Studies
   4.1 Sources of Uncertainty in External Dosimetry 21
   4.2 Sources of Uncertainty in Internal Dosimetry 27
   4.3 General Approach to Uncertainty Analysis Based on Dosimetric Parameters 32
   4.4 Numerical Models for Error Propagation in Uncertainty Analysis 35

5. Effects of Dose Uncertainties in Radioepidemiologic Risk Regressions 42

6. Methods to Adjust Risk Regressions
   6.1 Simulation Extrapolation 45
   6.2 Regression Calibration 49
   6.3 Likelihood-Based Methods and Monte Carlo Maximum Likelihood 52
   6.4 Bayesian Markov Chain Monte Carlo 56
   6.5 Two-Dimensional Monte Carlo with Bayesian Model Averaging 58
   6.6 Corrected Information Matrix 61
   6.7 Moment Reconstruction and Moment-Adjusted Imputation 63
   6.8 Summary 65

7. Summary and Recommendations 70
Appendix A. – Monte Carlo Simulation for Radiation Dose Confidence Interval with Negligible Shared Error

A.1 Python code
1. Executive Summary

1.1 Background

Absorbed dose from exposure of the human body to ionizing radiation cannot be directly measured and therefore all doses must be estimated based on indirect measurements and mathematical or biophysical modeling. Therefore, uncertainty in dosimetry is an inevitable concomitant of radiation health effects studies and has the potential to distort the results of a study in terms of the relationship between radiation dose and the studied outcome in dose-response models if statistical methods are not used to account for it. The purpose of this Commentary is to evaluate different, commonly used, statistical approaches for correcting for uncertainties in dosimetry estimates and to review the types of uncertainty to which they can be applied. Previous NCRP reports have looked at and characterized dosimetric uncertainty in external and internal exposures to ionizing radiation (NCRP 2007, 2009b), and have also looked at uncertainty in study results, but no NCRP document has thoroughly examined the variety of statistical methods to correct for dosimetric uncertainty. This Commentary aims to thoroughly investigate the variety of methods that can be used to account for dose uncertainty.

1.2 Characteristics of Dosimetric Uncertainty

Dosimetric uncertainty arises due to uncertainty in the parameters and structure of a dosimetry system to estimate radiation dose and measurement error in the data input to the dosimetry system. There are important distinctions regarding the type(s) of dosimetric uncertainty that apply in any study. Uncertainty refers to the limitations in our knowledge or understanding of a specific process, measurement, or parameter. This can be due to several factors, such as measurement error, data limitations, or a lack of knowledge about the underlying process. Most importantly, some uncertainties are random among individual members of a cohort, whereas others may be shared among subcohorts or the entire cohort. Uncertainties can either be of classical or Berkson type, as determined by the statistical relationship between true dose and the estimated dose. Some studies are based on dosimetry that has complex uncertainty, comprising a mixture of Berkson and classical error, particularly those incorporating internal exposure to radionuclides deposited in various tissues and organs of the body and for which doses are calculated using detailed biokinetic
models. Dosimetric uncertainty is reviewed here, although more complete reviews of methods of assessment of dosimetric uncertainty are provided in earlier NCRP reports (NCRP 2007, 2009b).

### 1.3 Effects of Dosimetric Uncertainty on Radioepidemiologic Studies

Dosimetric uncertainty unavoidably increases the uncertainty in estimating related health risks of exposure to radiation. This increase in the uncertainty in risk estimates cannot be reduced by statistical methods, but it is important to account for it, to correctly estimate the effect that dose uncertainty has on the uncertainty of risk estimates, including addressing such issues as shared uncertainty in dose. In addition to increasing the uncertainty of risk estimates, the other result of dosimetric uncertainty when it is of classical non-differential form may be to affect the slope of the dose-response curve, attenuating it either downward for linear dose response, or reducing its apparent curvature for a quadratic dose response (Carroll et al. 2006, van Smeden et al. 2019, Keogh et al. 2020). It is these latter effects on the slope or curvature of the dose response for which statistical corrections can be made. The statistical methods for correcting for radiation dose uncertainty in dose-response curves are reviewed here.

### 1.4 Statistical Methods to Correct for Dosimetric Uncertainty

This Commentary reviews the strengths and weaknesses of seven different statistical methods that are in the current literature and describes the types of uncertainty to which they can be applied. *Simulation Extrapolation* (SIMEX) is a conceptually straightforward method in which increasing amounts of randomly generated uncertainty are added to the dose estimates and the risk regression is repeatedly refitted, resulting in a relationship that is then back extrapolated to zero dose uncertainty. *Regression calibration* is a time-honored analytical method in which the risk is assessed by replacing the true dose in regression models by the expected value of the (unavailable) true dose estimated from the observed (uncertain) dose estimates based on knowledge or assumptions regarding the distribution of true dose and the statistical relationship between true and uncertain dose. *Likelihood-based methods*, in particular *Monte Carlo Maximum Likelihood* (MCML) methods, use multiple probability models for relationships like those between true and uncertain dose and between true dose and epidemiologic outcome, and are often estimated by
Monte Carlo methods making use of the complex dosimetric uncertainty models that have been developed in current studies. *Markov Chain Monte Carlo* (MCMC) is a Bayesian likelihood-based method that assumes prior distributions for the parameters including the radiation risk and then finds a posterior distribution by successive applications of Bayes’ theorem. It is solved by Monte Carlo methods as the solution is generally analytically intractable. *Two-dimensional Monte Carlo* (2D MC) with *Bayesian Model Averaging* (BMA) is a method that uses Monte Carlo sampling to generate multiple sets of estimated doses with both shared and unshared errors, hence the “2D,” and solves a risk regression for each set, then combines the results in an average weighted by their goodness of fit, hence the BMA. *Corrected Information Matrix* (CIM) is a method that uses a score statistic to correct the confidence intervals of the risk estimates for the shared Berkson uncertainty. *Moment Reconstruction* (MR) and *Moment Adjusted Imputation* (MAI) are methods that use the first two or more moments of the estimated distribution of true doses to adjust the observed uncertain doses to have the same moments.

### 1.5 Conclusions and Recommendations

Each method has advantages and disadvantages that need to be weighed in each individual application. Below is a summary of the recommendations included in and expanded on in this Commentary.

- A rigorous analysis of dosimetric uncertainty should be included in every study and statistical methods should be used to correct for the effect on the risk regression.
- It is found that some methods are quite general whereas others can only be applied to a particular type of uncertainty.
- Regression calibration, moment reconstruction, Monte Carlo Maximum Likelihood and Bayesian Markov Chain Monte Carlo are the most general, although Moment Reconstruction and Moment Adjusted Imputation are most suited to certain problems.
- Corrected Information Matrix and Simulation Extrapolation are less general but suitable for some situations, i.e., models with only Berkson uncertainty in the first case and only classical uncertainty in the second.
• Some methods such as Monte Carlo Maximum Likelihood and Bayesian Markov Chain Monte Carlo are very computationally intensive and may be prohibitively demanding of computer resources for large cohorts and complex models of dosimetric uncertainty.

• Two-dimensional Monte Carlo with Bayesian Model Averaging is quite general but has decisive statistical problems that need to be considered.

• A software repository should be created for some of these methods.

• More research, and working groups, should be devoted to developing methods for accounting for dosimetric uncertainty in combination with model uncertainty and outcome misclassification.
2. Introduction

There is direct evidence of human health risks at moderate and high levels of ionizing radiation dose (UNSCEAR 2008; IARC 2012), which is the type of radiation dose referred to in the paragraphs below. Although there are several studies yielding estimates of radiation risk at low doses (i.e., <100 mGy) for highly radiogenic cancers such as leukemia and thyroid cancer (Lubin et al. 2017; Little et al. 2018; Little et al. 2022a; Little et al. 2022b), for many cancer sites, as well as for outcomes other than cancer, it is necessary to assess risks via extrapolation from groups exposed to moderate and high levels of dose. These studies have been reviewed by NCRP (2018) and in a monograph produced by the National Cancer Institute (Daniels et al. 2020). A major source of uncertainty in radiation risk estimates arises from the extrapolation of risks at high doses and high dose rates to those at low doses and low dose rates. Crucial to the resolution of this uncertainty is statistical modeling of the dose-response relationship and implementation of both systematic and random dosimetric errors in dose-response models. Assessment of dosimetric uncertainty is vital for deriving unbiased risks from the studies of the Japanese atomic bomb survivors that are most often used to quantify population risks (ICRP 2007; UNSCEAR 2008; Ozasa et al. 2019).

Dosimetric uncertainty arises due to uncertainty in the parameters and structure of a dosimetry system to estimate radiation dose and measurement error in the data input to the dosimetry system. The problem of allowing for dosimetric uncertainty when estimating dose-response relationships has been the subject of much interest in biostatistics (Pierce et al. 1990; Pierce et al. 1992). It is well recognized that dosimetric uncertainty can alter the shape of this relationship substantially and hence the derived population risk estimates (Carroll et al. 2006).

In this Commentary, we focus on analytic methods to account for dosimetric uncertainty in regression models. We do not assess other analytic challenges in radiation epidemiologic studies, such as outcome misclassification, error in other covariates such as smoking, or related issues in study design. We survey assessments of quantification of uncertainties in dosimetry and go on from that to consider the effects of dosimetric uncertainties in risk regressions, and statistical methods to adjust risk regressions in radiation exposed groups for these uncertainties. The dosimetric uncertainty issues addressed in this Commentary are found in radioepidemiologic studies of exposure from many sources including occupational, environmental, and medical
exposure and including space radiation. Throughout the Commentary we provide examples from large and well-characterized epidemiologic cohorts to demonstrate applications of the statistical concepts presented; they are not intended to be exhaustive. Furthermore, while most of the examples pertain to ionizing radiation, the statistical methods reviewed are applicable to both ionizing and non-ionizing radiation. We end with conclusions on recommended methods to correct for dosimetric uncertainty.
3. Definitions of Uncertainty and Error Types

![Diagram showing the relationship between different types of uncertainty in radiation dosimetry.]

**Fig 3-1.** The relationship between different types of uncertainty in radiation dosimetry.

Uncertainties can be classified as representing either variability (representing variations of (possibly known) characteristics in the population), random error (representing unknown factors operating at an individual level) or bias (systematic error, representing unknown but fixed factors applying to whole groups) (NCRP 1996). Furthermore, the random errors can be classified as either shared or unshared, and each of these may also be classified as classical or Berkson form (NCRP 2007, 2009b). These are illustrated in Figure 3.1.

Uncertainty and variability are distinct but are related concepts. Uncertainty refers to the limitations in our knowledge or understanding of a specific process, measurement, or parameter. This can be due to several factors, such as measurement error, data limitations, or a lack of knowledge about the underlying process. For example, when assessing radiation doses due to the ingestion of contaminated food, the average number of calories may be estimated using unreliable methods such as survey data. Uncertainty is often quantified using confidence intervals, which
provide a range of values within which the true value is likely to lie with a specified level of confidence. Variability, on the other hand, refers to the natural differences or dispersion present in a dataset or a population. This can be due to inherent diversity in the data or the randomness of a particular process. Variability is a fundamental property of data and can be observed across various measurements or observations. For example, the caloric intake of a family may vary with age, sex, and socioeconomic status. Common measures of variability include range, variance, and standard deviation, which provide insight into the spread or dispersion of the data.

Uncertainty can be either random or systematic. Random uncertainty typically arises from statistical fluctuations that are inherent to any measurement of a particular quantity [i.e., repeated measurement under identical conditions produces a distribution of results that represents the operation of unknown processes that can be thought of as fundamentally probabilistic (i.e., random) in nature]. Systematic uncertainty is nonrandom in nature and is associated with biases in estimation of a particular quantity (i.e., a tendency for estimates to be greater or less than the true value). Systematic uncertainty can arise, for example, when a measurement device is improperly calibrated or there is an inherent and unknown bias in a scenario, model, data set, or other assumption. Typically, in random uncertainty, the difference between the true and estimated or measured value is assumed to have zero mean, whereas in systematic uncertainty the difference between true and estimated values is some nonzero (but possibly unknown) value, which will generally result in bias.

3.1 Classical or Berkson Error

When considering uncertainty in an estimate of a quantity, such as a dose or a parameter in a dose calculation model, it is important to consider the probabilistic nature of the uncertainty (i.e., of the difference between that estimate and the unknown true value of the quantity). In many settings, the estimate differs from the true value by an error, that is:

\[
\text{estimate} = \text{true value} + \text{measurement uncertainty}
\]

where: measurement uncertainty is a random variable with mean zero and is statistically independent of true value. This is the classical measurement error model, which arises naturally
in settings where observations are made on the individual units for which the quantities are defined, and each observation is subject to random perturbation due to such factors as instrument imprecision and recording uncertainty or misreport (Carroll et al. 2006). For example, among the Japanese atomic bomb survivors, the location of survivors at the time of the bombing may be subject to errors in recall (in interviews some years after the bombing) and recording of these. The uncertainties resulting from determination of dose from film badge readings in an occupational setting are often of classical type (Gilbert et al. 2006). It is important to recognize that classical measurement error leads to an observed variability of individual dose estimates that will be larger than the variability of true dose among individuals (Carroll et al. 2006). Inflation of the inter-individual variability of true dose caused by classical measurement error may bias the estimate of the epidemiologic dose-response function towards the null (i.e., the value of the slope of the dose response will be smaller than if the dose response were analyzed based on true values of dose) (Carroll et al. 2006; van Smeden et al. 2019; Keogh et al. 2020) and will increase the variance of the error in the dose response estimates (Carroll et al. 2006). In contrast, the Berkson error model, in which the true value varies from the estimate by an uncertainty that is random and independent of the estimate, can be written:

\[ \text{true value} = \text{estimate} + \text{individual deviation} \]

where: \textit{individual deviation} is a random variable (with mean zero) and is independent of \textit{estimate}. The Berkson model arises, for example, when the mean dose for a group of individuals is applied as a surrogate for the true dose to each individual belonging to that group, as for example is sometimes done in groups of underground miners (Allodji et al. 2012) or persons exposed to environmental sources of radiation (Masiuk et al. 2013). Among atomic bomb survivors, those for whom shielding histories were not collected, who were predominantly at locations beyond 1.5 km in Hiroshima or 2 km in Nagasaki, have their shielding calculated with transmission factors that are averages of those for whom there was no information on radiation shielding by buildings and adjacent structures (Cullings et al. 2006). Unlike the classical error model, Berkson errors will not introduce bias in the dose-response parameter estimates in a linear model but will result in inflation of the related error variance (Carroll et al. 2006; Zhang et al. 2017); however, bias is possible for models with nonlinear link functions (Carroll et al. 2006; Oraby et al. 2018).
It is useful to consider the distinction between classical and Berkson errors when evaluating the uncertainty distributions of the input data and model parameters in a dose reconstruction model. This will help ensure that the uncertainties or states of knowledge about the parameters are properly represented in the uncertainties of the calculated doses. Frequently, “environmental parameters” (generally those that are shared by all individuals across a particular dose realization) have grouping (Berkson) uncertainty structure, and those that are specific to the individual (if known) have a classical measurement uncertainty structure. However, generic models and assumptions are frequently used to fill in for lack of information about specific individuals (e.g., dose coefficients); in this case, the errors in the parameter can be considered to be Berkson in nature, provided that the true mean for the group of individuals concerned is well known.

3.2 Shared and Unshared Error

In dose reconstruction scenarios it is quite common that uncertainties are shared by the study subjects, and characterization of the accuracy and precision of dose estimates may be improved if shared uncertainties are quantified (UNSCEAR 2015). The uncertainty in a parameter value or component of dose is unshared if it has a unique value or distribution of values for each subject whose dose is estimated, statistically independent of the analogous quantity or ensemble in all other individuals, and shared if the parameter value or component of dose has a common value or distribution of values in a group of study subjects, so that the inter-subject correlation is 1. Some shared uncertainties are easy to identify; for example, the crew of a naval vessel present at an atmospheric nuclear weapons test at the Pacific Proving Grounds share the uncertainty of vessel position relative to surface zero of the weapon explosion. However, members of these crews do not necessarily share uncertainties related to their individual position in the ship above or below the weather deck or the waterline. Similarly, a ground unit present at a test at the Nevada Test Site (NTS) share uncertainties in weapon yield and position of the trench line from ground zero, but may not share the uncertainties in exposure from fallout on the ground from previous tests if portions of the unit took different routes during maneuvers from the trenches towards the blast site or exhibit areas. In occupational exposure scenarios, data available typically include dates of employment, job title, and duration of employment, as well as dosimeter reading and bioassay results. However, workers at the same site, and in the same building, with the same job title (e.g.,
chemical operator), may have had very different exposure conditions. Again, worker interviews and compilations of data about exact working conditions are useful in evaluating which uncertainties in the dose reconstruction are shared or unshared among which subjects. Generally speaking, some of the uncertainties in dosimeter readings or bioassay analysis results will be shared if the same on or off-site dosimetry service was used. All workers whose dose estimates depend on bioassay data will share the uncertainties arising from the application of the standard internal dosimetry models and reference individual parameters to these different individuals. Similar shared uncertainties arise for external dosimetry from issues of exposure geometry and unmonitored dose.

### 3.3 Other Types of Uncertainty

This Commentary focuses on dosimetric uncertainty in epidemiologic studies of radiation-exposed populations. However, additional sources of uncertainty, including outcome errors (or, misclassification) and statistical model uncertainty, also deserve attention but are out of scope for this Commentary.

There have been a few analyses of radiation-exposed groups in which uncertainty in the statistical model associating dose with outcomes has been taken into account. The risk models used to study radiation-exposed groups are often of two forms: excess relative risk (ERR) and excess absolute risk (EAR). The use of mixture models has sometimes been employed, in part as a way of making inference on the types of model (Muirhead and Darby 1987; Little et al. 1997). One can go further than this and attempt to take account of uncertainties in the various types of model. For example, Walsh and Schneider (2013) and others (Walsh and Kaiser 2011; Schöllnberger et al. 2012) used a method for assessing the relative weight to be attached to ERR and EAR model estimates. The method is based on the established statistical technique of multi-model inference (MMI) (Burnham and Anderson 2002; Claeskens and Hjort 2008). Although not explicitly Bayesian, MMI is somewhat related to Bayesian model averaging and related Bayesian techniques (Wang et al. 2012); these Bayesian methods have the advantage of assessing the parameter uncertainty distribution more thoroughly than MMI, albeit at somewhat greater computational cost. However, BMA and MMI have also been much criticized, and MMI in particular may not adequately take account of uncertainties. Little and Wakeford (2013) pointed
out that this method may be misleading because it uses purely statistical methods for weighting ERR and EAR models, making inference from just a single exposed group [e.g., the LSS in the case of Walsh and Schneider (2013)]. There is no reason why BMA and MMI should not take account of dosimetric error, for example via regression calibration, but this has yet to be done to the best of our knowledge.

Another source of error is outcome misclassification which arises when outcomes are incorrectly classified, resulting in bias toward or away from the null which leads to either under- or over-estimation of exposure-outcome associations. Examples of outcome misclassification include binary case-control outcomes for which a case is incorrectly classified as a control (i.e., false negative) or a control is incorrectly classified as a case (i.e., false positive). In radiation epidemiology, outcomes are often collected from administrative sources, such as employment or health registry records, that can be subject to errors, including incorrect cause of death on death certificates or incorrect type of primary tumor in cancer registries. Outcome misclassification has been accounted for in a few analyses. In particular, Sposto et al. (1992) used expectation-maximization (EM) methods to model outcome misclassification in the LSS, using autopsy data as a way of making adjustment for misclassification between cancer and noncancer endpoints. However, this analysis took no account of dosimetric error. French et al. (2020) quantified the potential impact of death-certificate inaccuracies on radiation risk estimates for liver cancer in the LSS via simulation, for which true-positive and false-negative rates were obtained from a previous study that compared death-certificate causes of death with those based on pathological review (Sharp et al. 2001). Many cohorts lack the validation data with which to correct for outcome misclassification.
4. Dosimetric Uncertainties in Radioepidemiologic Studies

In many assessments, the estimated quantity can be a discrepant estimate of the true value, which can itself be subject to possibly unknown variability, and the discrepancy can be positive (i.e., a tendency towards overestimation) or negative (i.e., a tendency towards underestimation). When the discrepancy results from some systematic discrepancy of some quantity from the true one, bias is said to occur. Bias can occur, for example, when a measurement technique yields systematically high or low results, is improperly calibrated, or is applied correctly but with a deficient device. In some circumstances a single sample or measurement can be considered an unbiased estimate of the mean of the distribution describing the measurement error. However, bias can occur if the measurement errors have a log-normal distribution and the measured value is the median rather than the mean of the log-normal distribution. Biases can arise from sources other than the measurement process itself. A significant and common source of bias is an incorrect assumption used in the calculation of dose. For example, a bias can occur when an inhalation dose is calculated assuming that the radionuclide inhaled was in a soluble form, when, in reality, the radionuclide was in an insoluble form.

Measurement errors in exposure assessment, and the dosimetric uncertainties that result from them are inherent in many radiation epidemiologic studies (NCRP 1996, 1997). Such errors arise partially from uncertainties in unknown parameters and model structure in complex dosimetry systems. Sources of uncertainty can be propagated through the exposure assessment to produce an estimate of uncertainty for the assessment of the final dosimetry endpoint. While dosimetric uncertainty is unavoidable, it is possible to characterize sources of uncertainty in the inputs of a dosimetric evaluation. Even the best personal dosimeters measure dose from external radiation at a point outside the body, not in the tissues, and have other uncertainties in addition to those in the relationship between dose at the dosimeter and dose in the tissues (NCRP 2018). Even the best measurements related to internally deposited radionuclides, such as bioassays, have considerable uncertainty in the estimates of dose to tissues that derive from the related estimates of tissue content of the radionuclide(s) of concern. Of course, many studies do not have sources of dose information such as personnel dosimeters and/or measurements such as bioassay on the entire cohort. This section provides an introduction to the complexities of uncertainties encountered in radiation epidemiology, placing emphasis on dosimetric uncertainty and its ramifications for analyzing
radiation-exposed populations (NCRP 1996, 1997). Undertaking a thorough uncertainty analysis, in other words an assessment of magnitudes of uncertainty and their correlations in various components of dose, is imperative for generating dependable risk assessments and facilitating well-informed public health decisions. It is vital to systematically pinpoint, quantify, and address diverse sources of uncertainties, including measurement inaccuracies, model uncertainties, individual exposure variabilities, and outcome misclassification, as well as potential dependencies among exposure assessment model parameters related to the distribution of dose-related quantities in the cohort and shared errors that may emerge from specific population subsets (ICRP 2010; NCRP 2012).

Various methods and techniques, such as Monte Carlo simulations and expert judgement, are utilized to address uncertainty in radiation epidemiology research (NCRP 2018). Monte Carlo simulations offer insight into the probability distribution of uncertainties and their influence on dose estimations (NCRP 1997), while expert judgement is vital to determine parameter uncertainties, particularly when data are limited or incomplete (ICRP 2009). By seeking input from experts in the field, researchers can acquire subjective probability distributions for uncertain parameters, which can subsequently be integrated into the analysis. These techniques empower dosimetrists to gain a deeper understanding of the probability distribution of uncertainties and their effects on dose estimates (NCRP 2009a, 2009b).

This section focuses on illustrating methods of characterizing and communicating dosimetric uncertainty from reconstructions through two exemplary case studies (ICRP 2004). So-called shared error can result when two individuals have an identical component of dose, or when there is a common parameter determining a component of dose in both persons. The first case study examines the determination of the dosimetric probability density function for individuals where there is negligible shared error (IAEA 2007). The second case study emphasizes the development of dose estimates for multiple realizations, which is a technique of particular importance when dealing with the expectation of substantial shared errors (NCRP 2009a; Simon et al. 2015). These instances underscore the significance of accounting for both shared and unshared errors in cohort studies, while also touching upon other types of uncertainties, such as outcome errors and statistical model uncertainty (EPA 1999, Pawel et al. 2007).

While additional sources of uncertainty are briefly discussed, a comprehensive treatment of these topics is beyond the scope of this section. The primary objective is to provide a thorough
examination of the diverse methods and techniques employed in confronting uncertainty in radiation epidemiology research and to demonstrate their practical applications in real-world situations (NCRP 2001). By exploring these case studies, the intention is to showcase the efficacy and versatility of Monte Carlo simulations and other techniques in addressing dosimetric uncertainty, enabling researchers to make more informed decisions and develop robust risk assessments that consider the intricate nature of uncertainties inherent in radiation exposure studies.

Quantitative uncertainty analysis usually requires that the state of knowledge about the uncertain components of the mathematical model be described by probability distributions. In the absence of cohort-specific data, these distributions are derived using available evidence on the model parameters, the model structure, and any dependencies that exist among model components, supplemented in some cases by expert judgment. Indeed, under most circumstances, expert judgment must be employed because, even when site-specific data exist, they are seldom complete. Probability distributions obtained using expert judgment are inevitably subjective. They represent the degrees of belief about what are probably fixed but unknown quantities, such as model parameters, correlations between these and model structure, often described by a mean or central estimate and covariance terms. Sometimes there is thought to be some true (but unknown) distribution of a certain model parameter, representing for example heterogeneity in the underlying population. Uncertainties resulting from such variability in the population and uncertainty due to lack of knowledge about fixed quantities should be evaluated separately.

When relevant cohort-specific data are unavailable, the most defensible method for obtaining subjective probability distributions is through the formal elicitation of expert judgment using a structured questionnaire or interview (Little et al. 1997a, 1997b). Formal elicitation is used to encode what is known about a model component. Because of the expense involved, exploratory analyses are recommended to first identify the model components of dominant importance to the overall estimate of uncertainty. Expert elicitation then focuses on refining the quantification of the state of knowledge for these components. In the absence of formal expert elicitation, important assessments should be assigned to at least two independent organizations, and the assessment should include resources for resolving discrepancies. This applies to uncertainties of all types, Berkson and classical.
4.1 Sources of Uncertainty in External Dosimetry

The reader is referred to NCRP Report No. 158 (NCRP 2007) and NCRP Report No. 178 (NCRP 2018) for a comprehensive investigation of the sources and extent of uncertainties associated with measurement techniques and estimation of organ doses from exposure to external radiation. Several practical case studies that illustrate the concepts have been discussed in these reports.

An individual can be exposed to radiation from sources outside the body by a variety of ionizing radiations such as x rays, gamma rays and beta particles, in a number of ways. Measurements can be made of the radiation field in air (or the energy absorbed in material such as a film badge or other type of dosimeter) resulting from these radiation sources. The measured quantity may be ionization or air kerma ($K_a$) or even the angular and energy distribution of the flux of a particular radiation at a point in space. The uncertainty in the measured quantity can usually be inferred from a combination of measurements and models. A model must be used to relate the particular measured quantity to the expected absorbed dose in tissue $D_T$. This estimated $D_T$ will always be uncertain, both due to the error in the underlying measurement itself and to the uncertainty in the model and model parameters used to relate this measured quantity to the absorbed dose $D_T$ in a body organ of a particular individual. To some extent, the relationship between a measured or specified quantity in air and $D_T$ can be estimated by measurements using a phantom that is a physical mockup of the human body. However, the uncertainty in the conversion to $D_T$ is generally only an estimate based on overall scientific knowledge and expert judgment. Because all measurements involve some degree of uncertainty and one cannot directly measure the absorbed dose to a human organ, all estimates of dose will be uncertain.

One common framework of estimating external dose to cohort members is first measuring the radiation field, then relating the characteristics of that field to organ doses. For instance, several area monitors may be used to estimate the energy and intensity of the photons (x rays or gamma rays) in a room with workers. Next, the information estimated about the photon flux incident on the workers can then be used to estimate their organ doses. In this dosimetry paradigm, there are uncertainties associated with both the characterizing of the radiation field, and the subsequent estimation of organ doses.
Radiation field measurements can be categorized as either area monitors or personal monitors. Furthermore, these monitors can be subdivided into devices which measure photon and neutron fields or doses (or beta doses). The major sources of uncertainty arise from imperfect knowledge of the radiation field for which the instrument was employed. Improper assumptions about the radiation field may lead to differences in energy response and angular response that are assumed for the area monitors or personnel dosimeters (ICRP 2010). Furthermore, other significant sources of measurement error may arise from readout bias, limited resolution, improper calibration procedure and small detector integration periods. Modern regulations place limits on detector inaccuracy at calibration, but these assurances cannot be assumed for dosimeter readings made earlier than the 1970s.

Even if the radiation field is properly characterized in terms of energy, particle type and orientation, substantial uncertainty may arise in the subsequent determination of organ doses. True dose to tissues and organs cannot directly be measured and are typically calculated using models and assumptions about the radiation fields. These models have parameters that cannot practically be known with certainty. For example, it is often important to know what fraction of time a worker’s body is oriented directly towards a source as opposed to standing in the same position, but facing away from the source. Improperly characterizing this single factor may cause an order of magnitude error when calculating the organ doses of specific organs for low-energy sources. Unique challenges arise depending on whether the organ dose estimates are estimated from either area measurements or personal dosimeters.

Area (field) measurements characterize the radiation field rather than the exposure of a specific individual. Under this definition, the measurements would typically be made in air, or in or on inanimate objects, but more importantly, they are not made in or on the body of an individual. Area measurements do not have the advantage of moving around with the individual, as does a personal monitor, but they have other advantages. For example, some area measurements can collect types of data that would be impractical as personal measurements (e.g., collection of gamma-spectral data by in situ gamma spectrometry). Furthermore, area measurement detectors have less restrictions on size and weight; therefore they can be manufactured with larger, more sensitive active volumes and would generally provide more precise, accurate and multimodal results relative to the smaller, lighter personal detectors. Area measurements can also provide data
over periods of times much longer than would be practical for personal measurements and, hence, increase precision of the measurements.

Personal-dosimeter measurements are dosimetry data that more closely represent the exposure of an individual than area measurements. They are typically made using badges worn on the body that are processed or “read out” after a wearing period, but purely solid state electronic dosimeters are also available, which have the advantages of instantaneous determination of dose and dose rate. Personal-dosimeter measurements have the advantage that the detector moves with the individual through the radiation field. Hence, the measurements reflect the individual’s exposure to the field regardless of the temporal and spatial variations of the field and the amount of time spent at different locations in the field. Personal-dosimetry measurements typically use different technologies than area measurements and may be less informative about the quality and nature of the radiation field than area measurements. In addition, the limits of detection of personal measurements may be higher (i.e., having lower sensitivity) than field measurements that can use longer counting (observation) times.

Unlike area measurements that can be generalized to any individual moving through the radiation field, personal measurements are less useful to generalize to others than the persons wearing them, except possibly for cases where it can be assumed that other persons moved in the field through the same locations for the same time periods. This may indeed be an acceptable assumption for certain situations (e.g., estimating the average dose for military troops traveling on foot). This was done, for example, for military personnel participating in early nuclear-weapons tests. In those cases, individuals were usually not issued individual dosimeters, but rather their radiation exposure was inferred from a limited number of cohort dosimeters issued to a few members of their unit (NRC 2003). Depending on the requirements of a dose assessment for individuals in a military unit, it may be necessary to acknowledge that not every member of such units would have traversed the exact same path through the radiation field. Whether the differences in dose and the related uncertainty for individuals taking slightly different paths is important for the dose assessment is not obvious. Lack of other necessary dosimetry information could easily introduce much greater uncertainty.

With the complexity of the absorbed dose calculation from personal measurements in mind, it is essential to consider the various sources of uncertainty when calculating organ dose. The complexity of the dose calculation from personal measurements is compounded by the fact that
the radiation field is often composed of a mixture of radiation types, such as beta-gamma and neutron-gamma, with an energy spectrum that is difficult to characterize. Furthermore, the angular distribution of the radiation field is often unknown, and the individual's position relative to the radiation field is uncertain, leading to the possibility of partial-body shielding. In addition, individual body size and depth of the organ of interest can vary, and the frequency and placement of the dosimeter on the body is not always known. All these factors contribute uncertainty to the absorbed dose to a particular organ or tissue, or a “whole body-equivalent dose,” from personal measurements.

The following sections provide “real-life” examples of assessments of uncertainties in dosimetry for external exposure in two large cohorts, the Japanese atomic bomb survivors and the 15-country nuclear worker study (predecessor of INWORKS).

4.1.1 Dosimetry of the Japanese Atomic Bomb Survivors

Extensive work has been done by large working groups to estimate the doses that Japanese atomic bomb survivors in Hiroshima and Nagasaki received by direct irradiation from the exploding bombs (Roesch 1987; Young and Kerr 2005). A series of dosimetry systems was constructed beginning in the 1950s, culminating with Dosimetry System 2002 (DS02) (Young and Kerr 2005; Cullings et al. 2006) which was amended in some respects, primarily the review and choice among multiple versions of survivor exposure records and revisions in terrain shielding, in DS02R1 (Cullings et al. 2017). DS02 and DS02R1 have been used by the binational Radiation Effects Research Foundation (RERF) in Hiroshima and Nagasaki and by many other researchers; Hiroshima University uses a somewhat less elaborate system called ABS93D and ABS16D (Hirota et al. 2021). The dosimetry systems calculate doses to individual survivors based on their reported locations at the times of the bombings, using detailed calculations of the radiation fluences at various distances along with structural shielding models of houses, and human phantoms to correct for body self-shielding (Cullings et al. 2006). Doses received after the bombings from neutron activation products in the survivors’ surroundings, or radioactive fallout, have been considered to be relatively minor and have not been included in the dosimetry to date by RERF (Roesch 1987), although some authors have suggested that such doses need to be considered (Tanaka et al. 2008; Endo et al. 2012). Recently, new work has been done and is ongoing by working groups including
RERF, University of Florida, National Institutes of Health and the Japanese Atomic Energy
Agency (JAEA) to construct a modern set of human phantoms (Griffin et al. 2019; Paulbeck et al.
2019) to replace the crude human phantoms based on primitive geometrical shapes that were
devised in the 1980s and adapted for Dosimetry System 1986 (Roesch 1987), reused in DS02, and
which are still in use (Sato et al. 2020).

A very detailed analysis of uncertainties in DS02 was made in Chapter 13 of the DS02 manual
(Young and Kerr 2005), starting with uncertainties in quantities such as the estimated heights of
burst of the two bombs, and many of the related errors are shared among all or subgroups of
survivors. However, none of the shared errors have been included in corrections made to date for
dosimetric errors in the radioepidemiologic studies of RERF (Cullings et al. 2017) or other
investigators. The uncertainties in survivor dosimetry have been considered to be dominated by
errors in survivors’ reported locations and to a lesser extent their reported shielding by buildings
(Jablon 1971; Cullings et al. 2017). A correction based on regression calibration (see section 6.2)
with an assumed 35 % geometric standard deviation (GSD) classical error (that is to say in which
\( \ln[\text{dose}] \) has classical error with standard deviation 0.35) has long been used at RERF and by many
somewhat different correction based on 20 % Berkson error and 40 % classical error has been
proposed (Pierce et al. 2008) but has not been adopted in other analyses. An analysis of
dyslipidemia in a small cohort of high-dose atomic bomb survivors with Multiple Indicators,
Multiple Causes (MIMIC) models using both classical and Berkson error resulted in an estimate
of 31 % classical error (Tekwe et al. 2014). The choice of an estimate for Berkson error is
complicated by the fact that there are survivors with and without shielding histories that give
details of the survivor’s location in or near a building, and that these histories were collected
mainly out to a distance of 2 km in Nagasaki, but only to about 1.6 km in Hiroshima, and not
exhaustively (Cullings et al. 2006). Survivors with a shielding history have their doses estimated
at RERF by a transmission factor that is calculated for one of a collection of positions in a model
house cluster that results in a modest Berkson error, whereas those without a shielding history have
their doses estimated by using structural shielding transmission factors that are averaged over large
groups such as all of the survivors with shielding histories who were 1) inside of houses, 2) outside
of houses, 3) completely in the open, or 4) in Nagasaki factories (Cullings et al. 2006), and hence
have considerably larger Berkson errors than those with shielding histories, except those relatively few survivors that were in the open.

### 4.1.2 IARC 15-Country Study and International Nuclear Worker Study

The dosimetry in the International Agency for Research on Cancer (IARC) 15-country study of nuclear workers, which covered various periods from the 1940s through 2000, has been intensively studied (Thierry-Chef et al. 2007). Errors in historical recorded doses have been identified and quantified based on a review of dosimetric practices and technologies in participating facilities. The main sources of errors in doses from “high-energy” photons (100–3000 keV) were identified as the response of dosimeters in workplace exposure conditions and historical calibration practices. Uncertainties related to dosimetry technology and radiation fields were quantified to derive period- and facility-specific estimates of bias and uncertainties in recorded doses. This was based on (1) an evaluation of predominant workplace radiation from measurement studies and dosimetry expert assessment and (2) an estimation of the energy and geometry response of dosimeters used historically in study facilities. Calibration factors, to quantify the errors relating to site- and calendar-period-specific dosimeter calibration practices were assessed. Site- and calendar-period-specific dosimetric factors were also assessed, to quantify errors concerned with the response of dosimeters to conditions of exposure (Thierry-Chef et al. 2007). These were combined into a site- and calendar-period-specific bias factor (Thierry-Chef et al. 2007). Coefficients were derived to convert recorded doses to $H_p(10)$ (dose at 10 mm below the body surface) and organ doses, taking into account different aspects of the calibration practices for dosimeters. A parametric, log-normal error model was developed to describe errors in doses as a function of facility and time period. Most errors, in particular the errors due to conditions of exposure, the energy response and angular response, and calibration and other administrative practices were thought to be mostly shared Berkson form (Thierry-Chef et al. 2007). The only classical errors in the dosimetry relate to dosimeter processing and reading, which were thought to be of unshared form, but also regarded as much more minor than the other errors (Thierry-Chef et al. 2007).

These dosimetric considerations mostly carry over directly to the International Nuclear Worker Study (INWORKS), the successor to the 15-Country study, which assessed only nuclear
workers in France, UK and USA, a subset of those in the earlier IARC study (Thierry-Chef et al. 2015). Changes made include updating the coefficients used to convert $H_p(10)$ to organ doses, reflecting updated International Commission on Radiological Protection (ICRP) models. Unlike the earlier study, assessments were made of neutron dose (Thierry-Chef et al. 2015).

### 4.2 Sources of Uncertainty in Internal Dosimetry

Each component that makes up the process of dose determination is also a source of uncertainty in the estimated dose. The sources of uncertainty can be divided into two major groups. One group includes the uncertainties introduced by bioassay or environmental measurements used to determine the activity of a radionuclide in the human body or in environmental media. The other group includes uncertainties in the parameter values and mathematical structure of the biokinetic and dosimetric models used in internal dosimetry. The reader is referred to NCRP Report No. 164 (NCRP 2009) and NCRP Report No. 178 (NCRP 2018) for a comprehensive investigation of the sources and extent of uncertainties associated with the estimation of organ doses from exposure to internally deposited radionuclides.

Because one cannot directly measure the absorbed dose to a human organ, all internal doses have to be estimated based on indirect measurements (e.g., bioassay data, environmental data) and based on mathematical models that simulate the bioaccumulation of the radionuclide and the transfer of energy between source and target regions. Given that models are only approximations of reality and they are built on the basis of sparse, and sometimes only partially-relevant data with varying levels of reliability (Leggett and Williams 1981), internal doses are uncertain.

The uncertainties in an internal dosimetry assessment may arise in several broad categories. These include uncertainties in measurements, in intakes, in biokinetic model structure and parameters, in dosimetric model parameters, and in the methods used to computationally calculate the organ doses.

The sources of uncertainties in direct bioassay measurements include calibration uncertainties that involve unknown distributions of activity in the subject, unknown thicknesses of overlying tissue, and other differences between the subject being measured and the calibration standard.

Uncertainties in indirect measurements arise in sampling, the critical issue of which being the representativeness of the sample; storage conditions, which include the duration of storage and
variable conditions during storage such as temperature and humidity; chemical analysis, the main
problem being chemical yield recovery; and counting, for which the critical issue is the calibration
of the measurement system.

Indirect bioassay measurements for activity levels approaching Minimum Detectable
Amounts (MDA), including all forms of error, are characterized by varied levels of measurement
uncertainties. Several examples illustrate this point:

- for $^3$H in urine via beta liquid scintillation, measurement uncertainties are in the range of
  10 % (NCRP 2018; Verrezen et al. 2008);
- for $^{90}$Sr in urine via chemical separation and beta counting, measurement uncertainties are
  in the range of 30 % (Eikenberg et al. 2006; NCRP 2018; Piraner et al. 2021);
- for actinides in urine via alpha spectrometry, measurement uncertainties are in the range
  of 30 % (Maxwell 2009; NCRP 2018); and
- for uranium in urine via inductively coupled plasma mass spectrometry (ICP-MS),
  measurement uncertainties are in the range of 10 % (Barwick et al. 1999; Becker 2005;
  NCRP 2018).

The parameters influencing the accuracy of in vivo monitoring were identified as:

- detector-to-body (or organ) distance;
- chest-wall thickness when counting inhaled $^{239}$Pu in the lungs;
- direction of scanning measurements versus the distribution of activity in the body;
- correspondence of the calibration phantom and the subject being measured; and
- distribution of activity in an organ being counted.

Radiation doses arising from inhalation occur in occupational settings and in
environmentally-exposed groups. An important source of uncertainty in estimation of doses from
inhalation is the physico-chemical form of the radionuclide. It is usually assumed that
radionuclides are attached to aerosol particles. However, this is not true in some important cases,
including inhalation of fallout particles in which the activity of refractory radionuclides is
distributed throughout the volume of particles that often are highly insoluble. If the radionuclide
is attached to aerosol particles, internal doses depend on the particle size distribution and on the
solubility of the radionuclide. In environmental exposure cases, particles sizes are normally
smaller, and soluble forms of radionuclides are more common than in the occupational setting.
Doses from radionuclides in gaseous forms depend on the chemical reactivity of the radionuclide.
Radioactive noble gases have little or no deposition in the respiratory tract. Iodine isotopes in air
can be found in a reactive chemical form [believed to be elemental iodine (I₂)], in rather inert
organic forms, and attached to aerosol particles of various sizes. Intakes of radioiodine, depending
to some extent on chemical form, result in uptake and deposition to the thyroid gland. Lack of
knowledge about the physical and chemical form of the radionuclide can introduce substantial
uncertainties in internal doses.

Intakes by ingestion are more common in the environmental setting than in the occupational
setting. In most cases of exposure in the environmental setting, bioassay measurements are not
available, and intakes by ingestion are derived from the radionuclide concentrations in the
consumed foodstuffs and the consumption rates of those foodstuffs. Knowledge of the physical
and chemical form of the radionuclide consumed in foodstuffs is less important than for the
inhalation pathway as the radionuclide is usually converted to an organic form during the
environmental processes involved from ground deposition to production of an agricultural
foodstuff.

Radionuclide concentrations in foodstuffs can be determined by direct measurements (as was
the case at some locations after the Chernobyl nuclear reactor accident) but more often, they are
calculated by means of environmental transfer models at the production stage. People consume a
large variety of foodstuffs, which have different origins and are prepared in various ways. The
consumption rates of foodstuffs vary widely according to the season, region of the country, degree
of urbanization, and type of foodstuff, among other factors. The foodstuffs are usually classified
into five categories (milk and milk products, meat, leafy vegetables, fruit and other vegetables,
and grain products), while water is considered separately.

Only rough estimates of the radionuclide concentrations in the critical foodstuffs and of the
 corresponding intakes can be obtained, because 1) there is a large variety of foodstuffs in each of
those categories; 2) consumption rates of individual foodstuffs vary widely; and 3) radionuclide
concentrations vary from one foodstuff to another and from one day to the next. Additionally,
intake estimates are complicated by factors such as the delay between production and consumption
for short-lived radionuclides, the method of preparation and its effect on gastrointestinal (GI) tract update, and the variation between foods consumed by various demographic subgroups.

Doses from internal emitters depend on the intake route of the radionuclides, on the amount that is transferred to blood, on the bioaccumulation and retention of the radionuclide in the various organs of the human body, and on the energy deposited per unit mass to any one organ by the radionuclide in the body. All these aspects represent complex processes that have been studied under certain conditions (e.g., human or animal studies, chronic or acute exposures, trace studies of the element of interest or of chemically similar elements). Based on usually limited experimental data, mathematical models have been designed to predict the transfer, bioaccumulation and retention of a radionuclide in various organs (i.e., biokinetic models), and to estimate the energy deposited per unit mass delivered to any one organ (i.e., dosimetric models).

The uncertainty in an individual’s dose may be understood as the distribution of possible true dose given whatever measurements have been made. However, in many cases, the evaluation of uncertainties requires analysis and interpretation of incomplete data and other complementary information, and it relies on professional judgment, a process that is largely subjective in nature. Thus, different analysts may produce different statements of dose and of uncertainty in dose for the same dose assessment endpoint. Since two analysts may produce different results, it is useful, at least for important and informative studies, to obtain assessments from several independent experts in the field, a process known as expert elicitation. One should note that expert judgment is needed at all stages of a dose assessment process, but expert judgment is a complement to, rather than a substitute for, other sources of scientific and technical information and data.

### 4.2.1 Mayak Nuclear Workers

Mayak is a facility in Russia (ex USSR) in which nuclear weapons were assembled, including plutonium cores. In the Mayak worker cohort (MWC), doses from external gamma rays have been calculated using various versions of the Mayak Worker Dosimetry System, in particular Mayak Worker Dosimetry System-2008 (MWDS-2008) (Vasilenko et al. 2007; Khokhryakov et al. 2013), MWDS-2013 (Napier 2014) and MWDS-2016 (Vostrotin et al. 2018); doses from internally deposited plutonium have also been estimated, using an updated dosimetry protocol (Birchall et al. 2017). Both MWDS-2013 and MWDS-2016 are probabilistic (with doses having a probability
distribution) and have been used to generate doses via a Monte Carlo Dosimetry System (MCDS) (Vostrotin et al. 2018) in which the probability distribution of doses is defined by successive ensembles of doses for all individuals in the cohort and all time periods. The MCDS produces estimates of possibly true doses by adding random variation to central estimates that are based on variables such as length of work in a particular plant during a particular time period. Annual doses from external gamma rays are available for each member of the study cohorts, also $H_p(10)$ personal dose equivalent for exposure to external gamma rays measured in Sievert (Sv). Absorbed doses from alpha radiation to lung, liver, bone surface and various other organs/tissues from internally deposited plutonium were estimated from measurements of plutonium in urine using biokinetic models of the behavior of plutonium in the body and dosimetry models (Birchall et al. 2017; Vostrotin et al. 2018). The current dosimetry system estimates doses for various parts of the respiratory tract, liver, bone surface, red bone marrow, gonads, kidneys, bladder, stomach, small intestine, colon and muscle, but does not provide dose estimates for blood vessels, heart or brain (Vostrotin et al. 2018). It is thought that the errors in dose in the MWDS are mostly of shared Berkson form (Zhang et al. 2017; Stram et al. 2021). As with other worker datasets there is likely to be a small component of unshared classical error due to film badge processing, but it is not clear whether this has been taken into account (Stram et al. 2021). Ensembles of MWDS-2018 lung dose estimates have been used to assess the effect of dose error for lung cancer in the Mayak workers using the CIM method (Stram et al. 2021).

### 4.2.2 Determination of Uncertainties of Radiation Exposure Assessment in the Wismut Cohort

The Wismut cohort was established in the 1990s and includes about 59,000 former male employees of the Wismut company (follow-up period 1946 – 2018) (Kreuzer et al. 2002, 2010). For each worker, a detailed working history was derived from the payrolls of the Wismut company. In 1998, the investigators developed a detailed job-exposure matrix (JEM) for estimating the exposure to radon progeny for employees of Wismut. This was further developed for scientific purposes in 2004 and implemented in a software program in 2005 (van Dillen et al. 2011; Küchenhoff et al. 2018; Deffner et al. 2019). The research project "Determination of uncertainties of radiation exposure assessment in the Wismut cohort, part 1" focused on the working conditions...
in uranium mining at the Wismut company in East Germany and procedures for estimating occupational exposure to radon and its progeny, as well as identifying potential sources of uncertainty and a preliminary evaluation of their possible relevance.

The sources and potential uncertainties in the exposure assessment are described and systematized in the report (Küchenhoff et al. 2018). The structure of uncertainties is complex because the multi-stage exposure assessment varies over time and depends on the working conditions and thus generates different types and sizes of errors. Errors in the exposure assessment may arise from generalization error (generalization of exposure measurements to a JEM with object-, calendar-year- and activity-specific exposure), assignment error (assignment of values in JEM to single workers) and estimation errors in all stages of the exposure estimation process. Generalization error and assignment error affect the exposure estimates of the whole cohort, while estimation error depends on the estimation approach used.

In a preliminary evaluation of the exposure uncertainties in the Wismut cohort study, two sources of uncertainty were considered particularly relevant: generalization error (e.g. use of average values for objects or shafts) and parameter uncertainties, specifically uncertainties in the determination of parameters (e.g. evaluation factors, activity factors). Other types of errors, for example resulting from evaluation of parameters by experts, human and technical errors in the measurement procedure, were judged to be much more minor (Deffner et al. 2019). These two sources can lead to exposure uncertainty when estimating radiation exposures for individuals within this cohort, which could have an impact on risk estimation if not properly accounted for. A comprehensive view of the type, distribution and structure of the uncertainties in exposure assessment was documented as an essential step towards the evaluation of the potential relevance of the uncertainties for the statistical analysis and the implementation of methods accounting for measurement error due to exposure to radon and progeny in the Wismut cohort.

4.3 General Approach to Uncertainty Analysis Based on Dosimetric Parameters

In most situations, uncertainty in the estimate of dose can be calculated from an estimate of uncertainty in each of the parameters used in the model equations. This approach is referred to as a "parameter uncertainty analysis" (IAEA 1989). The technique of parameter uncertainty analysis provides a quantitative method for estimating the uncertainty in the model result, assuming that
the structure of the model is an adequate representation of the real world and that correct values
for model parameters will produce correct results (NCRP 1996).

A parameter is uncertain if there is lack of knowledge about its true value. Parameter
uncertainty analysis should be composed of the following steps:

1. Define endpoint: Define the dose assessment endpoint.
   Example: In a radiation dose reconstruction study of a population living near a nuclear
   power plant, the assessment endpoint could be the average dose to the lung tissue from
   inhalation of radioactive particles. Another example might be a dose reconstruction study
   conducted to estimate the dose to lung tissue of individuals who were exposed to
   radioactive particles following the Chernobyl accident.

2. Identify uncertain parameters: Identify all potentially important uncertain parameters,
   including those that may have shared error, and add additional parameters if necessary to
   represent uncertainty in the model structure.
   Example: Uncertain parameters in the Chernobyl study could include the concentration
   and size distribution of radioactive particles released during the accident, the weather
   conditions at the time of the release, and the behavior of the exposed population (e.g.,
   amount of time spent indoors vs. outdoors, types of food consumed).

3. Determine the ranges of uncertain parameters: Specify the maximum conceivable range
   of values for each uncertain parameter, with respect to the assessment endpoint.
   Example: In the Chernobyl study, the concentration of the radioactive particles at an
   evaluation location could range from a 10 kBq/m³ to 200 kBq/m³, depending on the
distance from the release site. The size distribution of the particles could range from a 0.1
   µm to 2 µm. The weather conditions might range from calm and sunny to lightly turbulent
   affecting the transport of the airborne particles. The behavior of the exposed population
   could range from spending most of the time indoors to spending most of the time outdoors,
   affecting their exposure to the radioactive particles.
4. Determine probability distributions: Assign probability distributions that express the current state of knowledge about alternative values for each uncertain parameter range, taking into account both subjective and objective factors.

Example: In the Chernobyl study, a log-normal distribution might be used for the concentration of the radioactive particles based on measurements made in the years following the accident. A bimodal log-normal distribution might be used for the size distribution of the particles to account for different types of radionuclides. Meteorological data could be used to construct a probability distribution for relevant parameters such as wind speed and direction. Survey data or assumptions based on typical activity patterns could be used to determine the probability distribution for the behavior of the exposed population.

5. Identify interdependencies: Determine dependencies that are suspected to exist among parameters. In statistical evaluation, interdependencies can arise between two parameters when changes in one parameter influence changes in the other. This means that the values of the two parameters are not independent of each other, and understanding these interdependencies is essential for accurate analysis and interpretation of data. If enough information is available, then these dependencies can be characterized and used in the propagation of uncertainty to the estimates of organ absorbed dose.

Example: In the Chernobyl study, interdependencies might be identified between the concentration of the radioactive particles and the weather conditions, as higher wind speeds can cause more dispersion of the radioactive plume, resulting in a lower concentration at a given location. Interdependencies might also be identified between the size distribution of the particles and the behavior of the exposed population, as smaller particles can penetrate deeper into the lungs and affect the dose to specific regions of the lung tissue.

6. Propagate uncertainty: Use either analytical or numerical procedures to propagate the uncertainty in the model parameters to produce a probability distribution of model predictions, representing the resulting degrees of belief for the assessment endpoint. Monte Carlo simulations or other methods can be used to propagate the uncertainty in the
model parameters and produce a probability distribution of model predictions. It may be possible in some cases to express this uncertainty using a probability density function (PDF) representing the resulting degrees of belief for the assessment endpoint (see Fig. 3.2). In section 3.5 we explore numerical methods for error propagation in greater depth. Example: In the Chernobyl study, Monte Carlo simulations might be used to propagate the uncertainty in the model parameters and produce a probability distribution of thyroid doses for each member of the exposed population. A probability density function of thyroid doses for each member of the exposed population can be generated (Little et al. 2014; Likhtarov et al. 2014).

7. Examine and interpret results: Examine and interpret the results obtained from the uncertainty propagation process, considering the impact of shared error on the overall uncertainty of the dosimetric model. Example: In the Chernobyl study, the shape and characteristics of the probability distribution of thyroid tissue doses can be examined, as well as the sensitivity of the results to various uncertain parameters, including shared error due to interdependencies between the concentration of the radioactive particles and the weather conditions. Based on the results, recommendations can be made for further studies or actions to reduce uncertainty, such as improving the accuracy of the meteorological data used in the model or conducting additional surveys of the behavior of the exposed population.

4.4 Numerical Models for Error Propagation in Uncertainty Analysis

To overcome problems encountered with uncertainty propagation for complex models, various numerical methods can be used to perform an uncertainty analysis with the aid of a computer (NCRP 1996). Perhaps the most commonly applied numerical technique is Monte Carlo simulation. Other approaches include:

- Differential uncertainty analysis in which the partial derivatives of the response of the dose assessment model with respect to the parameters are used to estimate uncertainty.
The method is based upon the derivative and sensitivity data such as that calculated using direct or adjoint sensitivity analysis techniques (Puri and Ralescu 1983; Worley 1987);

- Monte Carlo analysis of simplifications of complex models (Labieniec et al. 1997; Simon 1999; Guyonnet 2003; de Wilde and Tian 2009); and

- nonprobabilistic methods, for example:
  o fuzzy sets and arithmetic (Puri et al. 1993; Cooke 2004; Baudrit et al. 2008)
  o possibility theory (Dubois et al. 2001)

### 4.4.1 Monte Carlo Analysis for Cohorts with Negligible Shared Error

During a Monte Carlo simulation, a value is selected at random from the probability distribution for each uncertain parameter, and the assessment model is run to produce a value for the prediction or assessment endpoint. This procedure is repeated for a specified number of iterations to provide a distribution of predicted values.

Monte Carlo analysis is usually performed using one of two random sampling processes: Simple Random Sampling (SRS) or Latin Hypercube Sampling (LHS) (Iman and Davenport 1980). In SRS, a random value is taken from the distribution specified for each uncertain model parameter, and a single estimate of the desired endpoint is calculated. This process is repeated for a specific number of samples or iterations (McKay et al. 1979). A set of $x$’s corresponding to each set of indices can be chosen by various methods, such as sampling at random from each interval or simply taking the interval midpoint. The result is an empirical approximation to the probability distribution of the model output or assessment endpoint.

SRS, however, is less efficient in many cases than LHS. To generate a sampling element by LHS of size $n$ from the variables $x_1, x_2, ..., x_N$ the range of each variable is divided into $n$ intervals of equal probability (i.e., $\frac{1}{n}$), so that $x_1$ has a set of $n$ interval midpoints $x_{N1}, x_{N2}, ..., x_{Nn}$. A single set of indices $(i_{11}, i_{12}, ..., i_{1N})$ is randomly selected, $1 \leq i_{1i} \leq n$ for all $i$, with an associated $N$-tuple of values $(x_{i_{11}}, x_{i_{12}}, ..., x_{i_{1N}})$. A second set of indices $(i_{21}, i_{22}, ..., i_{2N})$ is then chosen at random, so that $1 \leq i_{2i} \leq n$ and so that $i_{11} \neq i_{21}, i_{12} \neq i_{22}, ..., i_{1N} \neq i_{2N}$, leading to an $N$-tuple of values $(x_{i_{21}}, x_{i_{22}}, ..., x_{i_{2N}})$. This carries on in the same way, resulting in $n$ sets of indices $(i_{11}, i_{12}, ..., i_{1N}), (i_{21}, i_{22}, ..., i_{2N}), ..., (i_{n1}, i_{n2}, ..., i_{nN})$, all the indices are different for each
variable, and for each variable $x_m$ the set of values \(\{i_{1m}, i_{2m}, \ldots, i_{nm}\}\) is the same as the set \(\{1, 2, \ldots, n\}\) (ignoring order). The associated \textit{n-tuple of values} \((x_{i_{11}}, x_{i_{12}}, \ldots, x_{i_{1N}})\ldots, (x_{i_{n1}}, x_{i_{n2}}, \ldots, x_{i_{nN}})\) is the LHS; this is a sequence of \(n\) sample elements, where \(i\) is the value of variable \(X\) in sample element \(k\). The preceding process has the desirable property of producing a sample of relatively small size that covers the full range of each variable and has empirical cumulative distribution function close to that of the underlying distribution (Morio and Balesdent 2015).

The inputs required for Monte Carlo simulations are the probability distributions for each parameter. These distributions are obtained by applying professional judgment, measurements and results of computer simulations that reflect the current state of knowledge about each uncertain parameter, after extensive review of the available literature and data and evaluation of the relevancy of this information with respect to the assessment endpoint (see Section 4.3). Because professional judgment is used, the probability distributions are subjective. With the various input (subjective) distributions, the Monte Carlo simulation program then provides a subjective probability distribution of risk (or dose) from which subjective quantiles can be easily obtained. A demonstration of this technique follows.

Methods of representing the output of a Monte Carlo simulation include cumulative distribution functions and probability density plots. Monte Carlo analysis may be implemented by writing a sampling code or by utilizing one of several available software packages (see Appendix A). An example dose uncertainty calculation based on fish ingestion follows.

Situation. Let us assume that $^{137}\text{Cs}$ is continuously released to a nearby lake. Using the technique of uncertainty propagation (NCRP 1996), we calculate upper and lower quantiles that represent 95% of the distribution in dose to a maximally exposed individual resulting from eating fish from the lake. After review of the literature and available data, and after consultation with other experts, the (subjective) probability distributions shown in Table 4.1 can be derived for this problem. The contaminant concentration in fish ($C$) includes uptake of the contaminant from water by the fish; the contribution of drinking water to the individual's ingestion dose is assumed to be negligible in comparison to the dose from consumption of fish.
Table 4.1—Input parameter probability distributions assumed to propagate uncertainty in an individual’s dose estimate for fish ingestion example.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Central Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concentration in fish (C), Bq kg⁻¹</td>
<td>Log-uniform</td>
<td>$3.9 \times 10^{-2}$</td>
<td>$8.0 \times 10^{-1}$</td>
<td></td>
</tr>
<tr>
<td>Intake of fish by humans (I), kg d⁻¹</td>
<td>Log-uniform</td>
<td>$5.6 \times 10^{-3}$</td>
<td>$9.2 \times 10^{-2}$</td>
<td></td>
</tr>
<tr>
<td>Exposure frequency (EF), d y⁻¹</td>
<td>Constant</td>
<td></td>
<td></td>
<td>$3.5 \times 10^{-2}$</td>
</tr>
<tr>
<td>Exposure duration (ED), Y</td>
<td>Constant</td>
<td></td>
<td></td>
<td>$3.0 \times 10^{-1}$</td>
</tr>
<tr>
<td>Dose coefficient (DC), Sv Bq⁻¹</td>
<td>Log-triangular</td>
<td>$7.0 \times 10^{-9}$</td>
<td>$2.8 \times 10^{-8}$</td>
<td>$1.4 \times 10^{-8}$</td>
</tr>
</tbody>
</table>

The probability distribution function of the dose to a maximally exposed individual can be derived from the parameter probability distributions of fish concentration, fish intake, exposure frequency, exposure duration and the dose coefficient. The dose estimate is a simple product of these variables. A computer program can perform repeated statistical sampling and calculation to determine the distribution of doses using a Monte Carlo process. For this example, two log-uniform and a log-triangular distribution may be defined using the parameters listed in the table. A loop is then constructed to sample each distribution 100,000 times and calculate the dose for each sample. The empirical probability density function for dose is then calculated (see Python code in section A.1) and used to derive associated percentiles (Figure 4.2).
Fig. 4.2. Cesium-137 dose probability density function for fish ingestion example.

Distribution percentile values (2.5 %, 50 %, 97.5 %): $9.88 \times 10^{-5}$ Sv, $1.19 \times 10^{-4}$ Sv, $1.28 \times 10^{-4}$ Sv.

4.4.2 Monte Carlo Analysis for Cohorts with Non-negligible Shared Error

In the general approach to uncertainty analysis based on dosimetric parameters, it is essential to determine and account for dependencies that exist among parameter values. Sources of systematic error may only apply to specific population subgroups, making it difficult to assess shared errors. Commonly, numerous parameters may be used in dose assessment models, and because of the unique exposure characteristics of each subgroup, some parameter values may be shared within a subgroup or within the entire cohort. For example, an exposure assessment model may rely on parameters that depend on covariates, such as sex, sub-region and dietary practices. Therefore, errors in shared parameter estimates then result in shared errors in dose estimates for each of subgroup up to and including the entire cohort.

The mean and distribution quartiles, as might be determined by simple random sampling, does not account for shared error for individuals and would not be adequate to characterize these errors. This is because the potential for correlation between individuals is not considered.

If shared uncertainty is expected to be substantial, repeated draws from a Monte-Carlo procedure can be used to provide information on the shared and unshared uncertainty associated with the organ dose for each subject (Hofer 2007; Stram et al. 2015; Simon et al. 2015). A single individual’s dosimetric uncertainty can be characterized by the empirical probability density
function resulting from the Monte Carlo uncertainty analysis. Additionally, correlations between individuals’ doses can be determined by analyzing how the doses of a pair of individuals change across realizations. Thus, multiple realizations of the entire cohort’s dose (i.e., multiple sets of results from the Monte Carlo uncertainty model) can be used to account for shared uncertainties, unshared uncertainties and possible errors in the input data and the dosimetric models.

The goal of the multiple realizations (replications) dosimetry framework is to determine $M$ possible sets of doses for a cohort of $N$ subjects by appropriately sampling and assigning parameter values with defined uncertainties. Each parameter that contributes to the cohort member doses should be identified and, through expert judgement, should be assigned a parametric uncertainty distribution. For example, in a dose reconstruction involving the consumption of irrigated food, statements of uncertainty associated with vegetable, meat, milk and water ingestion rates, soil and water radionuclide concentrations, and dose coefficients should be stated. These uncertainty distribution specifications should be of a form where they can be readily sampled in a computer program. Additionally, for each parameter, it should be noted whether the uncertainties are shared or unshared. For example, individual ingestion rates may be defined as an unshared parameter, while soil radionuclide concentration may be defined as a shared parameter. Note that some parameter uncertainties may only be shared within a cohort subgroup (home gardeners, families with drinking wells, houses with enclosed basements, etc.). Once the state of knowledge of parameter uncertainties has been determined, the dose assessment proceeds by randomly sampling each parameter value. For each realization, shared parameters are only sampled once, and that single value is applied to all cohort members. On the other hand, unshared parameters are sampled for each cohort member. This can be achieved by two nested outer and inner loops in the sampling progress. Shared parameters are sampled in the outer loop and applied to all cohort members and unshared parameters are sampled in the inner loop and applied to each cohort member. The result is a dataset which contains a set of $M$ doses for each cohort member, for each evaluated organ, for each year of exposure. A small example with ten realizations, five cohort members, and two years of exposure for two organs is shown in Table 4.2a, b.
Table 4.2—Example data format of the multiple realization framework for five cohort members, two years of evaluation and ten realizations

a) Brain doses

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Organ</th>
<th>Year of Exposure</th>
<th>Organ Dose (Sv)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Realization 1</td>
<td>Realization 2</td>
</tr>
<tr>
<td>1</td>
<td>Brain</td>
<td>1952</td>
<td>0.45</td>
</tr>
<tr>
<td>2</td>
<td>Brain</td>
<td>1952</td>
<td>0.86</td>
</tr>
<tr>
<td>3</td>
<td>Brain</td>
<td>1952</td>
<td>1.28</td>
</tr>
<tr>
<td>4</td>
<td>Brain</td>
<td>1952</td>
<td>1.47</td>
</tr>
<tr>
<td>5</td>
<td>Brain</td>
<td>1952</td>
<td>1.31</td>
</tr>
<tr>
<td>1</td>
<td>Brain</td>
<td>1953</td>
<td>0.21</td>
</tr>
<tr>
<td>2</td>
<td>Brain</td>
<td>1953</td>
<td>0.30</td>
</tr>
<tr>
<td>3</td>
<td>Brain</td>
<td>1953</td>
<td>0.54</td>
</tr>
<tr>
<td>4</td>
<td>Brain</td>
<td>1953</td>
<td>0.55</td>
</tr>
<tr>
<td>5</td>
<td>Brain</td>
<td>1953</td>
<td>1.06</td>
</tr>
</tbody>
</table>

b) Lung doses

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Organ</th>
<th>Year of Exposure</th>
<th>Organ Dose (Sv)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Realization 1</td>
<td>Realization 2</td>
</tr>
<tr>
<td>1</td>
<td>Lungs</td>
<td>1952</td>
<td>1.76</td>
</tr>
<tr>
<td>2</td>
<td>Lungs</td>
<td>1952</td>
<td>2.08</td>
</tr>
<tr>
<td>3</td>
<td>Lungs</td>
<td>1952</td>
<td>3.35</td>
</tr>
<tr>
<td>4</td>
<td>Lungs</td>
<td>1952</td>
<td>1.73</td>
</tr>
<tr>
<td>5</td>
<td>Lungs</td>
<td>1952</td>
<td>3.04</td>
</tr>
<tr>
<td>1</td>
<td>Lungs</td>
<td>1953</td>
<td>3.11</td>
</tr>
<tr>
<td>2</td>
<td>Lungs</td>
<td>1953</td>
<td>3.05</td>
</tr>
<tr>
<td>3</td>
<td>Lungs</td>
<td>1953</td>
<td>3.01</td>
</tr>
<tr>
<td>4</td>
<td>Lungs</td>
<td>1953</td>
<td>0.98</td>
</tr>
<tr>
<td>5</td>
<td>Lungs</td>
<td>1953</td>
<td>3.53</td>
</tr>
</tbody>
</table>
5. Effects of Dose Uncertainties in Radioepidemiologic Risk Regressions

It is well recognized in the statistical literature that dosimetric uncertainty can have several effects, among the most serious of which is bias in estimates of regression parameters that associate radiation dose with risk of adverse health outcomes, as well as loss of statistical power to detect whether the association is statistically significant. So-called classical error, when nondifferential, will tend to result in bias of the coefficient estimates in dose toward the null (Carroll et al. 2006; van Smeden et al. 2019, Keogh et al. 2020) (Fig. 5-1). If there is curvature in the dose response, then classical error will generally result in reduction in the assessed upward curvature (Carroll et al. 2006) (Fig. 5-2). It will also result in increased uncertainty in the estimated coefficients (i.e., increased standard errors), and therefore lead to less precision and a reduction in statistical power (Carroll et al. 2006). However, as pointed out by Loken and Gelman (2017) an underpowered study with substantial classical error can yield bias away from the null. As Land (1980) has pointed out, if a low power study produces a significant trend estimate, the trend will likely be substantially upwardly biased.

So called Berkson error will not result in bias in a linear model, but will result in inflation of standard errors, and will therefore also result in loss of statistical power (Carroll et al. 2006). The only circumstance in which power is not lost is when tests are being conducted of the hypothesis that the dose coefficient is zero (Stram and Kopecky 2003).

The shape of the dose-response curve has profound implications for the extrapolation of risks at high doses and dose rates to those at low doses and dose rates. It is of course the possible risks arising from low dose and low dose rate exposure to ionizing radiation that are central to the setting of standards in radiation protection. For example, ICRP recommended a dose and dose-rate effectiveness factor (DDREF) of 2 based on animal and other radiobiological data, the evidence for curvilinearity in the cancer dose response in the atomic bomb survivor data and other epidemiologic data (ICRP 2007). Even in the absence of curvilinearity in the high dose rate dose-response curve there can be reduced effectiveness (i.e., lower excess risk) at lower dose rates, as has been observed in various sets of animal data (UNSCEAR 1993).

While the linear-quadratic dose response (with upward curvature) that is found for leukemia is perhaps the most often employed departure from linearity in analyses of the shape of the dose-response curve in radiation exposed groups (NRC 2006, UNSCEAR 2008), there are various other
possible shapes to the dose-response curve. For the class of tissue reaction (formerly deterministic) effects defined by the ICRP (2007, 2012), it is assumed that there is a practical threshold dose, below which there is less than 1 % incidence of effect. A linear threshold model has also been used to model cancer risk (Little and Muirhead 1996, 1997, 1998). Errors in dose can affect statistical inference on all these types of departure from linearity.

**Fig. 5-1.** Demonstration of the reduction in slope in a linear dose-response model fit to a hypothetical outcome (e.g., cancer risk) based on observed doses (depicted as turquoise points, with fitted dose response depicted as a turquoise line) versus true doses (depicted as red points, with fitted dose response depicted as a red line), for which classical errors are depicted by connecting black line segments.
Fig. 5-2. Demonstration of reduction of upward curvature for a linear-quadratic dose-response model fit to a hypothetical outcome (e.g., cancer risk) based on observed doses (depicted as turquoise points, with fitted dose response depicted as a turquoise line) versus true doses (depicted as red points, with fitted dose response depicted as a red line), for which classical errors are depicted by connecting black line segments.
6. Methods to Adjust Risk Regressions

In this section, we review currently available statistical methods for acknowledging and correcting for dosimetric uncertainties in statistical dose-response models that associate radiation dose with risk of adverse health outcomes to obtain unbiased and precise estimates for the dose-response curve. For each method, we describe the type of data to which the method can be applied, procedures for applying the method and its underlying assumptions, and a summary of the method’s strengths and limitations. We focus particularly on each method’s ability to accommodate shared uncertainties. We first introduce notation that is common across methods:

- \( Y_i \) = response variable for subject \( i, i = 1, ..., n \) (can be continuous or discrete/binary).
- \( X_i \) = true exposure covariate/latent variable, i.e., radiation dose, with standard deviation \( \sigma_x \).
- \( W_i \) = observed measure of the true exposure covariate/latent variable, with standard deviation \( \sigma_w \).
- \( U_i \) = unknown random error relating \( W_i \) to \( X_i \), with mean 0 unless otherwise specified and standard deviation \( \sigma_u \).
- Observed values of \( Y \) and \( W \) will be denoted with lowercase letters \( y \) and \( w \).
- \( \Theta \) = a vector of parameters relating \( Y \) to \( X \) in the dose-response model, e.g., a dose-response parameter (dose-response slope) \( \alpha \) and a vector of other ancillary parameters \( \beta \).
- \( Z_i = P \times 1 \) vector of error free covariates.

For simplicity, throughout we focus on linear dose-response models, but the functional form of the dose-response model can easily be modified depending on the context.

6.1 Simulation Extrapolation

The simulation extrapolation (SIMEX) (Cook and Stefanski 1994; Carroll et al. 2006) algorithm is suited for additive measurement error models (i.e., in which \( X_i = W_i + U_i \)) when reasonable estimates of the variance of the measurement error are available, e.g., reasonable values can be estimated from the observed data. SIMEX-based methods for measurement error correction are also applicable to conditions where the measurement error generating process can be simulated.
through Monte Carlo approaches (Carroll et al. 2006). Under this simulation-based mechanism to adjust the risk estimate for measurement error, additional amounts of measurement error are added to the observed measures of the true dose through simulation. Following the simulation steps, a trend (possibly with curvature) associated with the amount of measurement error is determined and SIMEX estimates are obtained via extrapolation of the trend. This simulation step is performed multiple times, while the extrapolation step involves extrapolating the established trends back to the hypothetical case of no measurement error.

### 6.1.1 Type of dose data to which SIMEX can be applied

In its original formulation (Cook and Stefanski 1994), the method is only applicable to situations where the error is of pure classical form and the magnitude of the error variance is known.

### 6.1.2 How SIMEX works

To apply the SIMEX approach, first assume that the forms of measurement error in the model for observed dose $W_i = X_i + U_i$ are additive where true dose $X_i$ and error $U_i$ are assumed to be mutually independent and also assume $U_i$ is independent of outcomes $Y_i$, $X_i$, and other covariates $Z_i$. One then adds to each observed dose a random independent error $\varepsilon_i$, with mean 0 and variance some small multiple $\lambda_1$ (generally less than 2 or so) of the known error variance, $\text{var} [\varepsilon_i] = \lambda \text{var} [U_i]$, and then fits the regression model in standard ways. This is done multiple times, with independent sets of errors, and the results averaged to yield an estimate of the parameters $\hat{\theta} (\lambda_1)$. This is then repeated for various other small values of $\lambda$ (generally less than 2 or so), $\lambda_1 < \lambda_2 < \cdots < \lambda_N$, to give $\theta (\lambda_1), \hat{\theta} (\lambda_2), \cdots, \hat{\theta} (\lambda_N)$. A curve $\hat{\theta}_{\text{fit}} (\lambda)$ is then fitted to these sets of parameters, as a function of $\lambda$, and extrapolated to $\lambda = -1$, to yield $\hat{\theta}_{\text{fit}} (-1)$, which represents the parameter estimates in the absence of dose error.

### 6.1.3 Assumptions
As noted above, the original formulation of SIMEX assumed pure classical error, and that the magnitude of the error variance was known. Generalizations allow for mixed Berkson and classical error (Misumi et al. 2018).

### 6.1.4 Procedure for applying SIMEX

We describe the SIMEX method in its original formulation (Cook and Stefanski 1994; Carroll et al. 2006), in which there are repeated measures of the observed dose \( W_i = (W_{ij}) \). The algorithm proceeds with the following steps:

1. Using the repeated measures on \( W_{ij}, j = 1, \ldots, J_i \), estimate the covariance matrix of the measurement error \( \hat{\Sigma}_{uu} \) as

\[
\hat{\Sigma}_{uu} = \frac{\sum_{i=1}^{J_i} (W_{ij} - \bar{W}_{i})(W_{ij} - \bar{W}_{i})^T}{\sum_{i=1}^{J_i}(J_i-1)}
\]

2. Identify a sequence of increasing positive small numbers such as \( \lambda = 0, \frac{1}{4}, \frac{1}{2}, \ldots, 2 \).

3. Simulate \( U_i \sim N(0, \sigma_U^2) \).

4. Obtain simulated values of the error prone covariate as

\[
W_i(\lambda) = W_i + \sqrt{\lambda} U_i \quad (6.1)
\]

5. At each value of \( \lambda \), maximize the likelihood associated with the model of interest or the regression analysis is performed by regressing the response on \( W_i(\lambda) \) adjusting for the error free covariates, \( Z_i \).

6. The simulation steps 4 and 5 are repeated \( S \) times and for the collection of all estimated coefficients in \( \hat{\theta}_s(\lambda) \) from the regression model associated with the \( s^{th} \) realization for \( s = 1, \ldots, S \), we calculate

\[
\hat{\theta}(\lambda) = \frac{1}{S} \sum_{s=1}^{S} \hat{\theta}_s(\lambda) \quad (6.2)
\]

7. The extrapolation step involves establishing an extrapolation function for the relationship between the estimated coefficients \( \hat{\theta}(\lambda) \) and \( \lambda \). The functional form for \( \hat{\theta}(\lambda) \) is purely empirical. Quite often a linear, linear-quadratic function or bilinear function is used (Cook & Stefanski 1994).

8. The extrapolation function evaluated at \( \lambda = -1 \) and \( \lambda = 0 \) yields the SIMEX and naive
estimators, respectively. The SIMEX estimators for the regression models $Y_i = \alpha X_i + \beta Z_i$ so that $\theta = (\alpha, \beta)$ are given by:

$$\hat{\beta}_{SIMEX} = \hat{\beta}(\lambda = -1)$$  
$$\hat{\gamma}_{p,SIMEX} = \hat{\gamma}_p(\lambda = -1), \; p = 1, \ldots, P$$

respectively. While the naïve estimators are expressed as:

$$\hat{\beta}_{Naive} = \hat{\beta}(\tau, \lambda = 0)$$  
$$\hat{\gamma}_{p,Naive} = \hat{\gamma}_p(\tau, \lambda = 0), \; p = 1, \ldots, P$$

Obtaining estimates of variance and confidence bounds for the risk estimates is difficult. Although the method requires having an estimate of the variance in the error of the dose estimate, estimating the variance of the risk estimate requires estimating the variance in the error of the back extrapolation to zero dose error. A resampling scheme such as bootstrap could be used to obtain such an estimate for SIMEX.

### 6.1.5 Strengths and limitations of SIMEX

As noted above, as originally formulated (Cook and Stefanski 1994) SIMEX requires that the errors are of pure classical form, with known error variance magnitude. The model has been generalized to allow for mixed classical and Berkson error (Misumi et al. 2018). It is computationally highly intensive so that depending on the number of bootstrap simulations and the complexity of the model, fits can take hours. R packages exist [e.g. SIMEX (Lederer et al. 2019)] to fit at least certain types of generalized linear model (McCullagh and Nelder 1989) although not the linear relative risk models in common use in radioepidemiologic studies. Quite apart from the computational difficulties, the method relies on a substantial extrapolation (from the given level of dose error to zero error), a jump that only provides an adequate risk estimate if the extrapolation function is specified correctly. Because of the computational overheads associated with this method, its use in any very large dataset may be impractical. Perhaps for this
reason and because of the limited types of error structure that can be handled it has been used only a few times to our knowledge, in analysis of the LSS data (Allodji et al. 2015; Misumi et al. 2018).

6.2 Regression Calibration

Regression calibration is a general method for dose error correction in which the occurrence of the true dose $X$ in regression equations is replaced by the conditional expectation of the true dose given the observed dose $W$. Regression calibration (Carroll et al. 2006) is commonly used in epidemiologic studies for bias corrections due to measurement errors. The method has been widely applied to measurement error correction in radiation epidemiology (Pierce et al. 1990, 1992; Little and Muirhead 1996, 1997, 1998, 2000; Little et al. 2008) and nutritional epidemiology (Spiegelman et al. 1997; Rosner and Gore 2001; Carroll et al. 2006; Bennett et al. 2017). Regression calibration can be applied in epidemiologic studies when the calibration equation $E(X|W,Z)$ is available or can be estimated from an ancillary study, which includes e.g., a gold standard for the true dose or replicates of the surrogate. It can also be applied in circumstances where no gold standard is available, as long as the errors are of classical form and a reliable estimate of the magnitude of classical error is available. It is an approximation-based approach for obtaining re-calibrated measures for the true unobserved exposures or covariates with the use of additional identifying data such as replicates or instrumental variables (Carroll et al. 2006; Batistatou and McNamee 2008; Yang et al. 2022). Thus, in the presence of additional identifying data for the classical additive measurement error models, specifically the magnitude of the error variance and the nominal dose distribution, the conditional distribution of the true covariate given the surrogate data may be approximated. Once the conditional distribution, $f(X_i|W_i,Z_i)$, is estimated, the true covariate $X_i$ is replaced with simulated data from this conditional distribution, in particular by the conditional expectation of the true dose conditional on the observed dose. The conditional expectations can be estimated via regression (of $X_i$ on $W_i$), using samples (possibly using multiple realizations) of the surrogate measures, $W_i$, and true dose $X_i$.

6.2.1 Type of dose data to which regression calibration can be applied
In principle the method can be applied to just about any type of data in which it is possible to estimate (or know from first principles) the conditional distribution of true dose given the nominal dose. In particular this is the case when the conditional distribution of nominal dose given the true dose is known (whether via some gold standard measurement set, or more theoretically estimated) and the true dose distribution is known (or can be estimated e.g. via deconvolution of the cumulative nominal dose distribution of \( W = \int f(W|X)g(X)\ dX, [0,W] \).

### 6.2.2 How regression calibration works

Using additional identifying information in the data, the conditional distribution of the true dose given the observed dose, \( f(X_i|W_i, Z_i) \), is estimated assuming the error has mean zero. Next, estimates of true doses \( E[X_i|W_i, Z_i] \) corresponding to the \( W_i \) may be calculated from \( f[X_i|W_i, Z_i] \). The calibrated estimates \( E[X_i|W_i, Z_i] \) are subsequently used in place of \( X_i \) in regression models.

### 6.2.3 Assumptions

Values of \( X \) and corresponding values of \( W \) or identifying information such as instrumental variables exist in the data to permit uniquely estimating \( \sigma_u^2 \), the variance of the measurement error or alternatively \( \sigma_x^2 \), the variance of the true dose. The error is assumed to have a mean of zero. Given an additive classical measurement error model, \( W_i = X_i + U_i \), where \( \sigma_w^2 \) is estimated from the model, estimation of either \( \sigma_x^2 \) or \( \sigma_u^2 \) is straightforward under the assumption that true dose is independent of the measurement error. This parametric approach to measurement error correction requires distributional assumptions for \( X_i \) and \( U_i \).

### 6.2.4 Procedure for applying regression calibration

Idea: The method is a two-stage estimation-based approach where the true doses are replaced by their approximated measures in the first stage. The plug-in replacement for true dose in the regression equations is the \( f(X_i|W_i, Z_i) \), conditional expectation \( E[X_i|W_i, Z_i] \). This conditional expectation may be calculated analytically given knowledge of the conditional density \( f(X_i|W_i, Z_i) \), and the magnitude of dose error, but may also be estimated via regressions based on
multiple samples of \((X_i|W_i, Z_i)\) which may also be produced via simulation. The second stage uses
the conditional expectations of true dose given the observed dose obtained from the first stage as
plug-in replacements in regression models. Likelihood-based model fits of the outcome (e.g. cancer) data as a function of the observed doses and other covariates can thereby be derived.

Obtaining estimates of variance and confidence bounds for the risk estimates is feasible because the method requires having estimates of the error variance or the variance of the true dose. It follows that one can estimate the error variance for \(E(X|W)\) and the resulting error variance for the risk estimates. See Table 6.2 for examples of studies with confidence bounds for risk estimates obtained by regression calibration.

A modified regression calibration method has been developed which is particularly suited to studies in which there is a substantial amount of shared error, and in which there may also be curvature in the true dose response (Little et al 2023). The method makes use of the 2\(^{nd}\) order terms in the Taylor expansion (with respect to true dose) of the likelihood. The second order terms in the Taylor expansion reflect both the effects of possible curvature in the true dose response and inter-individual correlations in the dose components (Little et al 2023). This method can be used in settings where there is a mixture of Berkson and classical error. For fits to synthetic datasets in which there is substantial upward curvature in the true dose response and varying (and sometimes substantial) amounts of classical and Berkson error, the coverage probabilities for the quadratic coefficient \(\beta\) were generally too low for the unadjusted and regression calibration methods, particularly for larger magnitudes of the Berkson error, whether this was shared or unshared (Little et al 2023). In contrast Monte Carlo maximum likelihood yields coverage probabilities for \(\beta\) that were uniformly too high (Little et al 2023). The modified regression calibration method yields coverage probabilities that are too low when shared and unshared Berkson errors are both large, although otherwise it performs well, and coverage is generally better than these other three methods. A notable feature is that for all methods apart from modified regression calibration the estimates of the quadratic coefficient \(\beta\) are substantially upwardly biased (Little et al 2023). The use of this approach requires knowledge about the magnitude of the measurement errors and some assumptions about the distribution of true dose given the observed data.

6.2.5 Strengths and limitations of regression calibration
A strength of the method is that it is usually simple to apply, and once the plug-in conditional expectations \( E[X_i|W_i, Z_i] \) have been derived (which can often be done in a spreadsheet) standard statistical software can be used. A limitation is that the method only takes account of first order error structure, so 2\(^{nd}\) and higher order terms in the Taylor expansion of the likelihood in terms of the error are not considered. However, as outlined by Little et al. (2023) the method may be amended to incorporate 2\(^{nd}\) order error structure, in particular to take account of shared error. The method works well when dose errors are small and the true dose response is linear or nearly so, but can give biased estimates if errors are large, or the true dose response is curved (Carroll et al. 2006; Little et al. 2023).

### 6.3 Likelihood-Based Methods and Monte Carlo Maximum Likelihood

Likelihood-based methods are an approach for accommodating errors in the observed dose variable—potentially through multiple realizations of the true dose—when performing point and interval estimation for regression parameters. They do this by integrating the expression for the likelihood as a function of the true dose vector, \( L((y_i), (Z_i), (X_i), (\beta), (\alpha)) \), over the distribution of true dose \( (X_i) \) conditional on the observed (nominal) dose, the dose response parameter \( \alpha \) and ancillary parameters \( \beta \). Inference on parameters of the resulting profile likelihood that is thereby derived is then performed in the usual way e.g., via maximum likelihood.

#### 6.3.1 Background on use of likelihood in regression

We present here an example, using aggregate (count) data, commonly encountered in radiation epidemiology. The methods described could be easily reformulated using individual outcome data. When data on cancer are described, typically the dependent variables \( y_i \) are count variables, recording the number of individuals with cancer (cases), or cancer deaths in group \( i \), which could have associated \( PY_i \) person years of follow-up. We may have various additional descriptive variables \( (Z_{ik})_{k=1}^{p} \) describing the mean attributes of individuals in that stratum, also the mean true dose \( X_i \). Then the model that could be fitted might be typically assumed in which the numbers of cases/deaths in stratum \( i \) is a Poisson random variable with mean given by:
\[ R_i(Z_i, X_i) = P Y_i \exp\left[\sum_{k=1}^{p} Z_{ik} \beta_k\right] [1 + \alpha X_i]. \] (6.7)

This corresponds to a Poisson rate process with intensity \( R(Z_i, X_i) / P Y_i \) cases/deaths per year. This is a linear relative risk model – the assumed baseline risk \( P Y_i \exp\left[\sum_{k=1}^{p} Z_{ik} \beta_k\right] [1 + \alpha X_i] \) is multiplied by the relative risk term \([1 + \alpha X_i] \). It is reasonable to expect that the expected numbers of cases/deaths will be proportional to the underlying cancer rate (per unit time) \( \exp\left[\sum_{k=1}^{p} Z_{ik} \beta_k\right] [1 + \alpha X_i] \) multiplied by the persons years of follow-up \( P Y_i \) in that stratum. The parameters \( \alpha, (\beta_k)_{k=1}^{p} \) are of course generally unknown. If \( y_i \) cases/deaths of cancer are recorded in that stratum then the probability of this observation, assuming that this variable has approximately Poisson distribution is:

\[
\exp[-R_i(Z_i, X_i)] \frac{(R_i(Z_i, X_i))^{y_i}}{y_i!} = \exp\left[-P Y_i \exp\left[\sum_{k=1}^{p} Z_{ik} \beta_k\right] [1 + \alpha X_i] \right] \frac{(P Y_i \exp\left[\sum_{k=1}^{p} Z_{ik} \beta_k\right] [1 + \alpha X_i])^{y_i}}{y_i!}. \] (6.8)

Putting together the observations over all strata we can write down the expression for the overall probability, or likelihood:

\[
L((y_i), (Z_i), (X_i), (\beta), \alpha) = \prod_{i=1}^{N} \exp[-R_i(Z_i, X_i)] \frac{(R_i(Z_i, X_i))^{y_i}}{y_i!} = \prod_{i=1}^{N} \exp\left[-P Y_i \exp\left[\sum_{k=1}^{p} Z_{ik} \beta_k\right] [1 + \alpha X_i] \right] \frac{(P Y_i \exp\left[\sum_{k=1}^{p} Z_{ik} \beta_k\right] [1 + \alpha X_i])^{y_i}}{y_i!}. \] (6.9)

This function can be maximized as a function of the unknown parameters \( \alpha \) and \((\beta_k)_{k=1}^{p}\). A common method of doing this is via the iteratively reweighted least squares algorithm (McCullagh and Nelder 1989). It should be noted that the dose vector \( (X_i) \) used in this expression is assumed to be the true dose. What we have described is simply one type of likelihood-based method, based on the Poisson linear relative risk models. Other types of likelihood, for example logistic likelihoods, are easily constructible, depending on the structure of the data.
6.3.2 Type of dose data to which likelihood-based methods and MCML can be applied

Likelihood-based methods can be used for conventional situations in which doses are measured with error (Carroll et al. 2006, Ch. 8). The related likelihoods can be evaluated by analytical methods (Carroll et al. 2006, Ch. 8) or by Monte Carlo methods (Carroll et al. 2006) (see below). In addition, likelihood-based methods can accommodate multiple realizations of radiation dose obtained from Monte Carlo dosimetry systems. These are situations in which there are no measured doses, but rather estimated distributions of true values for parameters that relate to doses, from which are sampled by Monte Carlo possible true values to generate vectors of possible true doses (Zhang al. 2017; Little et al. 2015) as discussed in more detail in section 3. For example, such a component might be the true dose rate for occupancy of a given area of environmental exposure in a given time period. When such parameters are shared among members of a cohort, the resulting vectors of results from the Monte Carlo dosimetry system are correlated for those members. Thus, both shared and unshared errors in doses can be acknowledged through potentially complex models for dosimetric uncertainty.

6.3.3 How likelihood-based methods and MCML work

Likelihood methods work by specifying a fully parametric model for each component of the data, including a model for the outcomes (as if true doses were observed) and a model for the errors (in the observed doses) (Shaw et al. 2020). The likelihood function is then identified by integrating the likelihood with respect to true dose \( X \), and the resultant profile likelihood function. When this integration and maximization is analytically intractable, Monte Carlo maximum likelihood can be used, which relies on replacing the integral of the likelihood \( L_\beta [y_i, (d_{ij})^{n}_{j=1}] \) for regression parameters \( \beta \) based on outcomes \( y_i \) and the true dose vector \( (d_{ij})^{n}_{j=1} \) [and possibly other covariates \( (x_{ik})^{n}_{k=1} \) conditional on the observed (nominal) dose vector \( (W_{ij})^{n}_{j=1} \) by its numerical average given a sample of \( N \) dose realizations of the dose vector \( (d_{ij})^{n}_{j=1} i = 1, ..., N \):

\[
E[L_\beta ((y_{ij})^{n}_{j=1}, (d_{ij})^{n}_{j=1}|(W_{ij})^{n}_{j=1})] \approx \frac{1}{N} \sum_{i=1}^{N} L_\beta [(y_{ij})^{n}_{j=1}, (d_{ij})^{n}_{j=1}|(W_{ij})^{n}_{j=1}] . \quad (6.10)
\]
That is, the likelihood is averaged across the distribution of the true doses given the observed doses. Integration of the resultant likelihood is performed using Monte Carlo sampling and then maximized with respect to the parameters $\beta$. Thus, Monte Carlo maximum likelihood can be used both to approximate maximum likelihood estimates and to construct approximations to likelihood-based confidence limits based on the profile log-likelihood (Stram and Kopecky 2003).

6.3.4 Assumptions

Maximum likelihood methods rely on parametric assumptions (i.e., a fully parametric model for each component of the data). The validity of the resulting estimates depends on the validity of the assumptions made for the various components of the likelihood.

6.3.5 Procedure for applying likelihood-based methods and MCML

Custom programs in standard statistical software (e.g., Stata, SAS) are available for conducting likelihood-based analyses for standard types of regression models (McCullagh and Nelder 1989). For doses on a continuous scale, standard numerical methods for integrating the likelihood (e.g., Gaussian quadrature) are typically applied. A simplified approach for Monte Carlo maximum likelihood can be implemented for Cox regression models based on partial likelihood methods, assuming homogeneity of the true doses given the observed doses across risk sets and that censoring is noninformative of the dose distribution (Stayner et al. 2007). However, this simplified approach is not applicable to Poisson or logistic regression models. Since this method generates a likelihood function for the risk estimates, confidence bounds can be estimated from the likelihood.

6.3.6 Strengths and limitations of likelihood-based methods and MCML

Assuming that the likelihood is correctly specified, likelihood-based methods should be more statistically efficient than those based only on first and second moments of the true dose distribution [e.g., moment reconstruction and moment-adjusted imputation as described in Schafer and Purdy (1996)]. Implementation of Monte Carlo maximum likelihood relies on specialized statistical software, often written in high-level languages such as Fortran or C/C++, and which
cannot be parallelized. Implementation can be highly computationally demanding, particularly when the null and alternative hypotheses are far apart. Computational demands grow when covariates impact the dose-response relationship (i.e., effect modification). Finally, Monte Carlo maximum likelihood suffers from the problem of the curse of dimensionality so that in integration over very high dimensional spaces of a rapidly varying function there may be insufficient points close to each other to well approximate the integral (Hastie et al. 2017). Sampling in a nonstandard way, for example using sequential Monte Carlo samplers (Dai et al. 2022), could ameliorate this problem.

6.4 Bayesian Markov Chain Monte Carlo

The Bayesian Markov Chain Monte Carlo (MCMC) approach for accommodating errors in the observed dose is based on the calculation of the joint distribution of the unknown parameters (via the posterior distribution). The model is constructed via conditional dependence relationships using four submodels, namely: (1) an outcome model that relates the outcome to the true dose and covariates; (2) a measurement model that relates the observed dose to the true dose; (3) an exposure model that relates the true dose to covariates; and (4) a prior distribution for all regression parameters (Richardson and Gilks 1993). There are distinct similarities with the maximum likelihood (MCML) approach discussed in section 5.3, which incorporates all but the last of these models.

6.4.1 Type of dose data to which Bayesian MCMC can be applied

Bayesian MCMC can be applied to data arising from several designs: data with a validation group assuming an error-free gold standard; data with several measuring instruments and/or longitudinal data (in the absence of a gold standard); and data with ancillary risk factor information (e.g., job exposure matrix). As such the Bayesian MCMC approach can be applied to all types of dosimetric data.

6.4.2 How Bayesian MCMC works
Bayesian estimation methods yield a posterior distribution for the parameters, which is obtained from an assumed likelihood function for the parameters given the data and an assumed prior distribution for the parameters, using Bayes theorem (Bayes and Price 1763). Derivation of the posterior distribution is generally analytically infeasible, and relies on the MCMC algorithm, specifically the Metropolis sampler, the Markov chain output from which is guaranteed to converge to the posterior distribution of the parameters.

### 6.4.3 Assumptions

As with maximum likelihood approaches, Bayesian MCMC relies on parametric assumptions (i.e., a fully parametric model for each component of the data: the outcome model, the measurement model, and the exposure model), which must be satisfied for estimation and inference regarding the regression parameters to be valid. In addition, the results can be sensitive to the choice of prior distributions. The assumed prior distributions are often chosen to be vague but must nevertheless be chosen with care; for example, a uniform distribution on the untransformed scale could yield different results from one which is uniform on a logarithmic scale. Also, the magnitude of the error variance is often specified, as a constant. The Bayesian MCMC approach can accommodate both Berkson and classical errors.

### 6.4.4 Procedure for applying Bayesian MCMC

Flexible and computationally efficient software exists to run Bayesian MCMC (e.g., Bayesian Inference Using Gibbs Sampling or “BUGS”). However, the method can be computationally demanding. A simplified two-step approach is available for fitting generalized relative risk models to grouped data in stratified person-year tables (Little et al. 2000; Bennett et al. 2004; Little et al. 2008; Little et al. 2020b), which in step 1 simulates means and variances of stratum-level doses from a Weibull distribution, and in step 2 obtains the posterior distribution using Gibbs sampling. Since this method generates a likelihood function for the risk estimates, confidence bounds can be estimated from the likelihood.

### 6.4.5 Strengths and limitations of Bayesian MCMC


A primary strength of Bayesian MCMC based on an assumed structure for conditional independence is that exposure uncertainty can be reflected in the variability of the parameters that quantify associations between exposure and outcomes. A further strength of the Bayesian approach is that the adjustment for measurement error is accomplished simultaneously to the estimation of the risk model and it can deal with complex error structures. Another key feature for the application of Bayesian MCMC as described in this section, as opposed to its use in 2-dimensional Monte Carlo with Bayesian model averaging (BMA), as described below, is that it avoids direct feedback in likelihood between dose and outcome by maintaining separate models for measurement, exposure, and outcome. However, it should be noted that in principle Bayesian MCMC methods do yield information on the true dose, which is part of the posterior distribution. However, in practice this feedback is often weak, and the 2-step method that has often been used (Little et al. 2000, Bennett et al. 2004, Little et al. 2008, Little et al. 2020b) in applications of Bayesian MCMC eliminates this possibility. In addition, while the convergence of model parameters may be guaranteed, the speed of convergence is not known, and in practice one can only rely on ad hoc tests of convergence.

### 6.5 Two-Dimensional Monte Carlo with Bayesian Model Averaging

Two-Dimensional Monte Carlo (2DMC) with BMA has been derived to make use of multiple realizations of dose produced via a MCDS. The BMA refers to the fact that the likelihood-weighted averaging of different risk estimates is done with results obtained from Bayesian MCMC solutions using different sets of MC dose results, i.e., “Bayesian” refers to the method of obtaining the results being averaged and not the method of averaging. The MCDS generates a number of sets of dose scenarios, one of which is assumed to be the “true” one. Health endpoint data are then combined with these using BMA to make inferences on the “true” scenario and thereby infer the dose coefficients (Kwon et al. 2016).

#### 6.5.1 Type of dose data to which 2DMC with BMA can be applied

This method is applicable to any setting where samples of doses are produced by a MCDS. A MCDS produces estimates of possibly true doses by adding random variation to central estimates
that are based on variables such as length of residence in a particular area during a particular time period or working in a particular area during a particular time period. The use of “two-dimensional” refers to the fact that the MCDS adds both individual variations and shared variations that are applied to all subjects or groups of subjects.

**6.5.2 How 2DMC with BMA works**

Ensembles of doses \((X_{ijk})^{N}_{j=1,k=1}\) are produced for all individuals for many realizations \(i, 1 \leq i \leq M\) of the MCDS. As noted by Stram et al. (Simon et al. 2021) the method is not described in sufficient detail to facilitate implementation. However, unlike other uses of MCDS it is assumed that only one of the dose scenarios \(i\), and therefore one of the sets of dose realizations \((X_{ijk})^{N}_{j=1,k=1}\) is the correct one. A dose-response function, for example the linear relative risk model \(R_i = f(Z_i)[1 + \alpha X_i]\), is then used to fit to each of these multiple realizations using Bayesian MCMC and the results are combined in a likelihood-weighted average, which is referred to by the authors as BMA (Hoeting et al. 1999). This reweights the scenarios depending on the goodness of fit (Kwon et al. 2016), and the ensuing Bayesian MCMC model fits are used to derive uncertainty intervals on the parameters of interest, for example the ERR \(\alpha\).

**6.5.3 Assumptions**

A key assumption is that one of scenarios represents the “true” set of doses, or at any rate is a very good approximation to the “true” set of doses. Essentially this method therefore assumes something like a combination of functional (that is to say where the unknown parameters are assumed fixed) and structural (that is to say where the unknown parameters are assumed to be drawn from some probability distribution) approaches – there are assumed to be random errors in the data, but certain parameters are assumed fixed (but unknown).

**6.5.4 Procedure for applying 2DMC with BMA**

The main work is the production of ensembles of doses \((X_{ijk})^{N}_{j=1,k=1}\) for all individuals for a large number of realizations \(i, 1 \leq i \leq M\). A dose-response function is specified, for example the
linear relative risk model $R_i = f(Z_i)[1 + \alpha X_i]$, is then used to fit to these multiple realizations using BMA (Hoeting et al. 1999). The model fitting is performed first followed by the model averaging. The posterior sample of the risk model parameters, for example the ERR $\alpha$, is produced using Bayesian MCMC, where Bayesian refers to the method of obtaining the results being averaged and not the method of averaging. That is, for each ensemble of MC-generated doses, a risk estimate is produced by Bayesian MCMC; then these risk estimates are combined in a likelihood-weighted average. Since this method generates a likelihood function for the risk estimates, confidence bounds can be estimated from the likelihood.

6.5.5 Strengths and limitations of 2DMC with BMA

In principle this method is reasonably easy to apply once one has an MCDS. However, as with all Bayesian methods it can be computationally cumbersome, and chain convergence difficult to determine. BMA is a likelihood-based method, and as such it requires separate modeling of outcomes and errors, as stated above in the section for likelihood-based methods (Carroll et al. 2006; Shaw, et al. 2020). However, it lacks these separate models. There are reasons to suppose that the method will produce substantially upwardly biased estimates of risk, as discussed by Stram et al. (Simon et al. 2021), because of circularity in the Bayesian weighting that is applied, also that the coverage may be poor. That is, the Bayesian weighting that is applied can favor the sets of Monte Carlo doses that produce the highest likelihoods in combination with the set of outcomes that actually occurred, with no accounting for the variability in outcomes that would be associated with a proper outcome model. However, this has been disputed by Kwon et al. (2016) who report that “When the estimated doses contain relatively small amounts of uncertainty, the Bayesian method using multiple a-priori plausible draws of dose vectors gave similar results to the conventional regression-based methods of dose–response analysis. However, when large and complex mixtures of shared and unshared uncertainties are present, the Bayesian method using multiple dose vectors had significantly lower relative bias than conventional regression-based risk analysis.” The observation that BMA was less biased than a naïve regression with no correction for dose error does not mean that it is unbiased. The implementation of the methodology presently relies on proprietary software, using MATLAB, and has not, to the best of our knowledge, been employed outside the group that developed it. Another substantial problem with the method is the
use of BMA, reflecting general criticism made of this class of models in the literature. BMA is a method that was developed to make inferences about a parameter using a set of models that differ, for example, in the other parameters that are included in each model. Its use with a set of identical models that differ only in the values that are used for the predictor variables is a questionable invention. BMA is only known to converge to the “true” model when one of the candidate models is really the true one. When this is not the case model convergence characteristics are less well understood. As with all Bayesian methods the choice of prior is critical. The priors used in 2DMC are flat - all scenarios are equally weighted.

6.6 Corrected Information Matrix

When dose errors can be assumed to be of shared Berkson form (also when Berkson errors are completely unshared) then theoretical corrections to the Fisher information matrix (Cox and Hinkley 1974) can be developed, which will result in expansion of the confidence intervals on the regression parameters (Zhang et al. 2017). The corrected information matrix (CIM) method only adjusts the confidence intervals. The central risk estimate is unchanged.

6.6.1 Type of dose data to which CIM can be applied

This method is applicable to any setting where samples of doses are produced by an MCDS, and in which errors are assumed to be of pure Berkson form.

6.6.2 How CIM works

An extensive calculation is required, described in more detail in the Appendix to Little et al. (2020). The score statistic (Cox and Hinkley 1974) and their covariance are derived, via knowledge of the form of the likelihood, and the empirical covariances (from the sample produced by the MCDS) of the doses. This is used to derive the asymptotic covariance of the model parameters via one step of the Fisher scoring algorithm (alternatively the Newton algorithm) (Little et al. 2020a). This yields the asymptotic covariance of the parameters, and this is used to derive the inflation to the CI.
6.6.3 Assumptions

A key assumption is that the dose errors which underly the production of the scenarios are of pure Berkson form, and that the ensemble of realizations is a sample from the true distribution of doses.

6.6.4 Procedure for applying CIM

The first step is the production of ensembles of doses \( (X_{ijk})_{j=1}^{n_j} \) for all individuals for a large number of realizations \( i, 1 \leq i \leq M \). The Berkson errors are generated using a model with both shared and unshared errors that are both multiplicative and additive (SUMA model). A dose-response function is specified, for example the linear relative risk model \( R_i = f(Z_i)[1 + aX_i] \).

The score statistic is then derived and its covariance estimated via the empirical covariance of the dose estimates. The asymptotic covariance is then derived, and this is used to compute the inflation in the CI of the model parameters. Further details are given in the Appendix to Little et al. (2020).

Software has been written in Python (Stram et al. 2021) and R (Little et al. 2020a) for application to specific problems. The software would have to be adapted for each specific dataset and risk model.

6.6.5 Strengths and limitations of CIM

Arguably the assumptions underlying the CIM method, that all dose simulations are samples from the true dose, may be unlikely, but this assumption is arguably less implausible than that made for 2DMC with BMA, which assumes that one realization is true. The method requires moderately extensive calculation, which can take on the order of an hour on a fast single threaded 64-bit workstation for a dataset of the size of the US Radiologic Technologists (~70,000 individuals) (Little et al. 2020a). The method appears to be well adapted to analysis of the Mayak data (Stram et al. 2021), where there is a substantial amount of shared error. In the USRT cataract data the amount of shared error is small, and the method yields largely trivial adjustments to CI (Little et al. 2020a).
6.7 Moment Reconstruction and Moment-Adjusted Imputation

Moment Reconstruction (MR) (Freedman et al. 2004; Freedman et al. 2008; Shaw et al. 2020) and Moment-Adjusted Imputation (MAI) (Thomas et al. 2011, 2013; Shaw et al. 2020) are methods to correct for error in the exposure variable by adjusting the observed exposure so that it has similar moments as the true exposure and hence the same distribution. MR uses only the first two moments, whereas MAI uses more, such as the first four moments.

6.7.1 Type of dose data to which MR and MAI can be applied

These methods are applicable to measured exposures, not MC generated ones, although of course the average or some other summary of MC generated values could be used.

6.7.2 How MR and MAI work

The method creates an adjusted exposure observation \( X_M(W, Y) \) or \( X_M(W, Y, Z) \) that has the same distribution as the true exposure \( X \), where \( W \) is the observed exposure, \( Y \) is the outcome, and \( Z \) represents covariates if any are included in the regression. The moments of \( X \) are estimated from a validation dataset containing observations of \( X, Y, \) and \( Z \), and are used in combination with the corresponding moments of \( W \) to adjust an observation of \( W \) to create \( X_M \) (Shaw et al. 2020). For example, in MR for a classical error model,

\[
X_M(W, Y, Z) = E(W|Y, Z) + G(W - E(W|Y, Z))
\] (6.11)

where \( G = \frac{\text{var}(X|Y, Z)^{\frac{1}{2}}}{\text{var}(W|Y, Z)^{\frac{1}{2}}} \).

MR and MAI have been applied to Berkson error and combinations of classical and Berkson error (Potgieter et al. 2016).
6.7.3 Assumptions

An error model involving classical or Berkson error or a combination of the two can be assumed. If the moments of $X$ must be assumed rather than being estimated from a validation sample, then these serious additional assumptions apply. There are cases in the literature in which a value for the variance of $U$ has been assumed rather than estimated from empirical data, such as the assumption of 35% classical error in the work on the Japanese atomic bomb survivors by Pierce et al. (1990), although this was used for regression calibration rather than MR. On the other hand, assuming values for the 3rd and 4th moments of $X$ seems much less feasible.

6.7.4 Procedure for applying MR and MAI

The main work is to estimate the required joint moments of $(X,Y)$ or $(X,Y,Z)$ as well as $(W,Y)$ or $(W,Y,Z)$ and use them to calculate the MR or MAI values of $X_M(W,Y)$ or $X_M(W,Y,Z)$ using related ratios, then substitute the $X_M$ into the regression in place of the $W$. There will be extra variability in the regression estimates due to error in estimating the moment-related parameters. This can be estimated by bootstrap methods i.e., randomly resampling subsets of the variables used in those estimates (Shaw et al. 2020).

6.7.5 Strengths and limitations of MR and MAI

Perhaps the main strength of MR and MAI is their breadth of applicability. Another advantage of MR and MAI is that they are easy to implement. They can be applied to logistic regression in addition to linear regression. They can even be applied in the case of differential error, i.e., where the error $U$ depends on the outcome $Y$ (Shaw et al. 2020), because the moments used to construct $X_M$ are estimated conditional on $Y$. A major limitation, of course, is the need for information on the moments of the true exposure $X$, with the issue that validation datasets containing $X$ are almost never available in radioepidemiology. A further limitation is their applicability for only a few types of regression models. A very natural comparison to make is with regression calibration, and interestingly MR and MAI are sometimes superior to regression calibration and other times inferior.
(Shaw et al. 2020). The STRATOS\textsuperscript{1} paper gives an example where MR is superior to regression calibration, which they say is because $Y$ is very strongly related to $X$ and $X$ is available in a validation dataset comprising a large part of the cohort (Shaw et al. 2020).

### 6.8 Summary

Table 5.1 summarizes the methods discussed in this section, with their advantages and shortcomings. It can be seen that perhaps no method is without disadvantages. Some methods, for example regression calibration (whether 1\textsuperscript{st} or 2\textsuperscript{nd} order) and moment reconstruction methods are relatively straightforward, are not computationally cumbersome and unlike many other methods do not rely on an MCDS. Table 5.2 (adapted from Wu et al. 2019) illustrates the range of adjustments for dose error seen in a few recent studies. Although in some cases the effect of adjustment is to increase risk, this is not always the case.

\begin{table}[h!]
\centering
\caption{Comparison of Methods}
\begin{tabular}{|c|c|}
\hline
Method & Advantage & Disadvantage \\
\hline
Regression Calibration (1\textsuperscript{st} order) & Easy to use & Requires strong assumptions \\
\hline
Regression Calibration (2\textsuperscript{nd} order) & More accurate & More complex \\
\hline
Moment Reconstruction & Simple & Requires validation data \\
\hline
\end{tabular}
\end{table}

\begin{table}[h!]
\centering
\caption{Examples of Dose Error Adjustments}
\begin{tabular}{|c|c|}
\hline
Study & Adjusted Risk \\
\hline
Wu et al. 2019 & Increase \%
\hline
\end{tabular}
\end{table}

\textsuperscript{1} www.stratos-initiative.org
Table 5.1—Summary of strengths and limitations of each method of dose error correction.

<table>
<thead>
<tr>
<th>Method</th>
<th>Classical/Berkson error</th>
<th>Useful for shared error</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation extrapolation (SIMEX)</td>
<td>Classical</td>
<td>Yes</td>
<td>Software exists in R to fit this class of models (although not using the standard linear relative risk models).</td>
<td>Only works in datasets with pure classical error, although model has been generalized to allow for mixed classical and Berkson error. Requires precise magnitude of dose error to be known. Computationally cumbersome. Method relies on substantial extrapolation (from given level of dose error in the data to zero error), a jump that only provides an adequate risk estimate if the extrapolation function is correctly specified.</td>
</tr>
<tr>
<td>Regression calibration 1st order</td>
<td>Classical &amp; Berkson (trivial for Berkson)</td>
<td>No</td>
<td>Relatively simple to apply Conventional statistical software can be used once conditional expectations have been derived.</td>
<td>Only works when errors are nondifferential (see glossary). Breaks down if dose error is large, or when there is curvature in the dose response. Does not account for correlations in dose or other higher order aspects of the dose distribution.</td>
</tr>
<tr>
<td>Modified regression calibration (2nd order)</td>
<td>Classical &amp; Berkson</td>
<td>Yes</td>
<td>Takes into account 2nd order terms in dose error, in particular intraindividual correlations between dose. Works better than 1st order regression calibration when dose response is curved or when dose error is large.</td>
<td>Requires special software be written e.g. in R.</td>
</tr>
<tr>
<td>Likelihood based methods and Monte Carlo maximum likelihood (MCML)</td>
<td>Classical &amp; Berkson</td>
<td>Yes</td>
<td>Takes full account of dose error distribution (in particular shared error). More efficient statistically than methods based only on first and second moments of the true dose distribution (e.g., MC and MAI).</td>
<td>Relies on Monte Carlo dosimetry system (MCDS). Requires specially written software using high level language (Fortran, C++). Computationally cumbersome. Problems when the underlying distribution is of high dimension – curse of dimensionality.</td>
</tr>
<tr>
<td>Bayesian Markov Chain Monte Carlo (MCMC)</td>
<td>Classical &amp; Berkson</td>
<td>Yes</td>
<td>Takes full account of dose error distribution (in particular shared error). Adjustment for measurement error is accomplished simultaneously with the estimation of the risk model. Approach can deal with complex error structures.</td>
<td>Computationally cumbersome. Choice of prior distributions may be critical. Speed of convergence is not known, and only ad hoc tests are available to inform when convergence has not taken place.</td>
</tr>
</tbody>
</table>
Models can be fitted using specialist software (WinBUGS, JAGS etc)

<table>
<thead>
<tr>
<th>Method</th>
<th>Error Type</th>
<th>Prior Requirements</th>
<th>Easy to Apply</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two-dimensional Monte Carlo (2DMC) with Bayesian Model Averaging (BMA)</td>
<td>Classical &amp; Berkson</td>
<td>Yes</td>
<td>Relatively easy to apply.</td>
<td>Relies on MCDS. Likely to produce upwardly biased estimates of trend. Computationally cumbersome. Coverage is likely to be poor. Choice of prior distributions may be critical. Convergence to “true” model is not guaranteed.</td>
</tr>
<tr>
<td>Corrected information matrix (CIM)</td>
<td>Berkson</td>
<td>Yes</td>
<td>Relatively fast to fit when dataset is of moderate size.</td>
<td>Relies on MCDS. Only relevant to data in which the errors are of pure Berkson form. Requires specially written software (e.g. in R, Python).</td>
</tr>
<tr>
<td>Moment reconstruction (MR) and Moment-adjusted imputation (MAI)</td>
<td>Classical &amp; Berkson (but trivial for Berkson)</td>
<td>No</td>
<td>Relatively simple to apply. Can be applied to situations with differential measurement error. Can yield less biased and more efficient results (than regression calibration) when dose errors are nondifferential, or when dose response is nonlinear.</td>
<td>Requires information on the moments of true exposure and validation datasets are rarely available. In some cases results are less efficient than for regression calibration.</td>
</tr>
</tbody>
</table>
### Table 5.2—Examples of results from radiation epidemiologic studies with and without adjusting for dose uncertainties, adapted from Wu et al. (2019).

<table>
<thead>
<tr>
<th>Study</th>
<th>Exposure</th>
<th>Outcome</th>
<th>Error-correction Method</th>
<th>Measures of Association</th>
<th>Unadjusted risk effect with 95 % CI^a</th>
<th>Adjusted risk effect with 95 % CI^a</th>
<th>Percent change after adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pierce et al. (1990)</td>
<td>External gamma and neutron dose</td>
<td>All cancer except leukemia, Leukemia</td>
<td>Regression calibration (35 % errors, 0-4 Sv)</td>
<td>Excess relative risk/Gy</td>
<td>NA</td>
<td>NA</td>
<td>+9.0 %</td>
</tr>
<tr>
<td>Little et al. (2014)</td>
<td>131-I thyroid dose (Ukraine)</td>
<td>Thyroid cancer</td>
<td>Regression calibration</td>
<td>Excess odds ratio/Gy</td>
<td>5.38 (1.86, 21.01)</td>
<td>4.78 (1.64, 19.69)</td>
<td>-11.2 %</td>
</tr>
<tr>
<td>Little et al. (2015)</td>
<td>131-I thyroid dose (Belarus)</td>
<td>Thyroid cancer</td>
<td>Regression calibration</td>
<td>Excess odds ratio/Gy</td>
<td>1.51 (0.53, 3.86)</td>
<td>1.31 (0.47, 3.31)</td>
<td>-13 %</td>
</tr>
<tr>
<td>Little et al. (2015)</td>
<td>131-I thyroid dose (Belarus)</td>
<td>Thyroid cancer</td>
<td>MCML</td>
<td>Excess odds ratio/Gy</td>
<td>1.51 (0.53, 3.86)</td>
<td>1.48 (0.53, 3.87)</td>
<td>-2 %</td>
</tr>
<tr>
<td>Land et al. (2015)</td>
<td>External thyroid dose</td>
<td>Thyroid nodules</td>
<td>BMA</td>
<td>Excess relative risk/Gy</td>
<td>4.16 (0.54, 7.77)</td>
<td>1.47 (4.1 × 10^-5, 3.74)</td>
<td>-64.7 %</td>
</tr>
<tr>
<td>Kesminiene et al. (2008)</td>
<td>Gamma-ray bone marrow dose</td>
<td>Hematological malignancies</td>
<td>MCML</td>
<td>Relative risk at 100 mGy</td>
<td>0.60 (-0.02, 2.35)^*</td>
<td>0.60 (-0.01, 2.58)^*</td>
<td>0 %</td>
</tr>
<tr>
<td>Stayner et al. (2007)</td>
<td>Gamma-ray whole-body dose</td>
<td>All cancer mortality (excluding leukemia)</td>
<td>MCML</td>
<td>Excess relative risk/Sv</td>
<td>5.38 (0.54, 12.58)^*</td>
<td>4.82 (0.41, 13.31)^*</td>
<td>-10.4 %</td>
</tr>
<tr>
<td>Allodji et al. (2015)</td>
<td>Gamma-ray colon dose</td>
<td>Solid cancer deaths</td>
<td>Regression calibration</td>
<td>Excess relative risk/Gy</td>
<td>0.432 (0.352, 0.512)</td>
<td>0.463 (0.377, 0.549)</td>
<td>+7.2 %</td>
</tr>
<tr>
<td>Allodji et al. (2015)</td>
<td>Gamma-ray bone marrow dose</td>
<td>Leukemia deaths</td>
<td>Regression calibration</td>
<td>Excess relative risk/Gy</td>
<td>3.858 (2.696, 5.020)</td>
<td>4.142 (2.895, 5.389)</td>
<td>+7.4 %</td>
</tr>
</tbody>
</table>

^CI = confidence interval  
*90 % confidence intervals  
**95 % Bayesian credible intervals
7. Summary and Recommendations

This Commentary reviews the different types of uncertainty and error related to radiation dose estimation and provides an overview of dosimetric uncertainties in exposure studies and their effects on radioepidemiologic risk regressions. The Commentary also reviews different methods for correcting risk estimates for the effect of dose measurement error, in particular regression calibration, Monte Carlo maximum likelihood (MCML), simulation extrapolation (SIMEX), Bayesian Markov Chain Monte Carlo (MCMC), two-dimensional Monte Carlo with Bayesian Model Averaging (2DMC with BMA), corrected information matrix (CIM) and moment reconstruction (MR). Each method has advantages and disadvantages that need to be weighed in each individual application. Based on this review the following recommendations are made:

Recommendation: Epidemiologists, statisticians and dosimetry teams should collaborate to include a rigorous and comprehensive analysis of dosimetric uncertainty in every radioepidemiologic study.

A recommended framework for evaluating uncertainty involves defining the dose assessment endpoint, identifying uncertain parameters, determining the ranges of uncertain parameters, assigning probability density functions to each uncertain parameter, identifying interdependencies and shared error, propagating uncertainty using analytical or numerical procedures, and analyzing and interpreting the results.

Recommendation: A method of adjusting for dosimetric uncertainty in the risk regression should be included in analyses of epidemiologic data.

- Regression calibration, MCML, and MCMC are quite general and valid methods, although MCML and Bayesian MCMC in particular can be computationally burdensome.

The most general methods are regression calibration, MCML, and Bayesian MCMC. Regression calibration is a very useful method, which can usually be easily applied but is useful only in situations with nondifferential measurement error. In the form usually used it only accounts
for the first order expansion in the error distribution, and will therefore break down when errors are very large; it also does not take account of inter-individual correlations between dose errors (i.e. shared errors) and is less suited to problems where there may be highly nonlinear dose response. However, regression calibration can be generalized to take account of second order terms in the error distribution, in particular shared errors, and in this form would also be better suited to cases where the dose response was nonlinear or there were inter-individual correlations in dose. MCML and Bayesian MCMC are more general methods, that can take full account of the error distribution, in particular of both classical and Berkson error, but they are computationally exceptionally cumbersome. MCML requires that special code be written in some high-level language (e.g. Fortran, C++). When large datasets are being analyzed using MCML or Bayesian MCMC consideration should be given to restricting the analysis to a carefully chosen subset of the most important uncertainties. Bayesian MCMC suffers from the problem inherent in most Bayesian methods, that convergence of the sampled chain is impossible to prove. The results may also depend on the choice of prior distributions, even when these are chosen to be vague.

- The CIM and SIMEX methods are suitable for particular applications but are limited in the problems to which they can be applied, and the CIM method is only a way of obtaining adjusted confidence intervals, not central estimates. (When errors are of Berkson type for a linear model the central estimate is not expected to change.)

Less general are the CIM method and SIMEX, and in some sense MR. The CIM method provides only a confidence interval adjusted for shared errors; a different method must be used to obtain central estimates. Furthermore, it can only be applied when the errors are of pure Berkson form, and in which the correlations between individuals can be estimated, for example using a MCDS. It requires that special code (in R or Python) be written. SIMEX is really suited only to problems with pure classical error and in which the magnitude of dose error is estimated. SIMEX is also quite computationally cumbersome. Nevertheless, for problems in which these types of error (pure Berkson, pure classical) are present, the SIMEX and the CIM methods can be useful. MR is particularly suited to problems in which there is thought to be a nonlinear dose response, or in which errors are nondifferential.
• The 2DMC method with BMA potentially has substantial limitations.

2DMC *per se* is just a way of saying that MC is used to generate both shared and unshared errors in the same application and is quite general. The use of 2DMC with BMA, however, involves generating sets of dose estimates with 2DMC, obtaining risk estimates for each set of doses using some conventional method such as Poisson regression, and then averaging those risk estimates using BMA. The 2DMC method with BMA has decisive statistical problems, in particular possible upward bias and lack of coverage, and since it relies on BMA the underlying assumption (that one of the models is the true one) appears to be implausible. As with Bayesian MCMC chain convergence is impossible to prove and only ad hoc tests to determine nonconvergence are available.

Recommendation: Create a repository of statistical software code for implementing some of these custom methods, specifically MCML, CIM, Bayesian MCMC and 2DMC with BMA.

Readily available, open-source statistical software facilitates the broader usage and application of advanced statistical methods while enhancing the transparency and reproducibility of data analyses and scientific reports. A central repository of software for implementing the statistical methods reviewed in this Commentary would provide researchers with improved access to these specialized methods. Possible repositories include GitHub or Open Science. Another alternative is an annotated library of the sort provided by the Comprehensive Epidemiologic Data Resource (CEDR) (DOE 2022) which is a U.S. Department of Energy (DOE) supported electronic database comprised of health studies of DOE contract workers and environmental studies of areas surrounding DOE facilities. CEDR is maintained by the Oak Ridge Institute for Science and Education (ORISE).

Recommendation: Devote research funding for comparative studies of existing and novel methods to evaluate and incorporate dose uncertainty, outcome misclassification, model uncertainty and their combined impact in epidemiologic studies. Establishment of an interdisciplinary working group to address these topics is encouraged.
In summary, the Commentary is based on a review of the current state of scientific literature regarding statistical methods to account for dose uncertainties in radiation dose-response analyses. The areas where substantial gaps in this literature exist are indicated, particularly regarding the performance of MCML, CIM, Bayesian MCMC, and 2DMC with BMA, both in general settings and in specific epidemiologic studies. Note that the Commentary does not consider methods for accommodating model uncertainty or outcome misclassification, nor the potential benefit of accounting for exposure and outcome errors that are correlated with one another, for which methods have been developed and applied in other scientific contexts (Shaw et al. 2021). Therefore, the final recommendation is for comparative studies of existing and novel methods be conducted to further guide and refine our understanding of the advantages and drawbacks of the methods. Research funding should be dedicated to such efforts, possibly through the DOE, and guided by an interdisciplinary working group of dosimetrists, statisticians, and epidemiologists.
Appendix A. — A Sample Monte Carlo Simulation to Determine Uncertainty in Radiation Dose Under Conditions of Negligible Shared Error

This appendix provides an explanation of the Monte Carlo simulation method used to determine the quantiles around an individual’s radiation dose, as presented in the main text. We describe the process in detail to ensure that readers who are not familiar with Python programming can still follow along and understand the underlying concepts. Accurately determining an individual’s radiation dose is a cornerstone of radioepidemiologic studies. Dosimetric uncertainty can profoundly affect the results of such studies, making the characterization of this uncertainty crucial. The Monte Carlo method is a computational technique that relies on random sampling to approximate numerical results. In the context of radiation dose assessment, it is particularly useful for estimating the uncertainty in dose calculations due to the inherent variability in parameters like radionuclide concentrations, intake rates, and dose coefficients. A range of methods for adjusting risk regressions and rectifying dosimetric uncertainty are examined in the main text, encompassing Simulation Extrapolation (SIMEX), Regression Calibration, Likelihood-Based Methods, and Monte Carlo Maximum Likelihood (MCML), among others. These methods all require information about the individual and shared error associated with the dose of cohort members. An example of a Monte Carlo method to characterize dosimetric uncertainty is detailed in this appendix. This simulation method allows us to generate a probability density of potential radiation doses, from which we can derive quantiles for an individual’s exposure. The following code demonstrates computational aspects of the simulation by providing a method and potentially a template for those wishing to employ this method in their own investigations. The code includes ample documentation so that programmers who are unfamiliar with python syntax can follow along. This example code generates a probability density of potential radiation doses, from which one can derive quantiles for radiation exposure.

Libraries and Modules

We begin by importing the necessary libraries and modules for our simulation:

- NumPy: A library for numerical computations in Python, which we will use for generating random samples and performing mathematical operations.
- Matplotlib: A plotting library for creating graphs and visualizations.
• Seaborn: An advanced data visualization library that builds on Matplotlib, which we will use to create a Kernel Density Estimation (KDE) plot of the dose distribution.

Custom Density Functions

Next, we define two custom distribution functions, loguniform and logtriangular, which will be used to generate random samples for the simulation:

• loguniform: This function generates random samples from a log-uniform distribution, which is characterized by a uniform distribution in the logarithmic space. This distribution is commonly used for modeling parameters that span multiple orders of magnitude.

• logtriangular: This function generates random samples from a log-triangular distribution, which is characterized by a triangular distribution in the logarithmic space. This distribution is useful for modeling parameters with a known central value and a range of variability.

Probability Density Functions and Constants

In this section, we define the probability density for the input parameters and the constants involved in the dose calculation:

• Radionuclide concentrations ($I$): A log-uniform distribution of radionuclide concentrations in fish, measured in Bq/kg.

• Daily intake of fish ($I$): A log-uniform distribution of the daily intake of fish, measured in kg/day.

• Dose conversion coefficient (DCC): A log-triangular distribution of the dose conversion coefficient, measured in Sv per Bq.

• Exposure Frequency (EF): A constant representing the number of days per year that an individual consumes fish, measured in days/year.

• Exposure Duration (ED): A constant representing the number of years that an individual is exposed to the radionuclide, measured in years.
Monte Carlo Simulation

We proceed by performing the Monte Carlo simulation:

- Generate random samples for each of the probability density functions (C, I, and DCC) using the custom functions defined earlier.
- Calculate the product of the probability density functions (C * I * DCC) and the constants EF and ED, to produce one realization of an individual’s radiation dose.
- Repeat steps 1 and 2 for a large number of iterations (e.g., 100,000) to generate the empirical distribution of potential doses.

Visualization

Finally, we create a KDE plot of the dose distribution using the Seaborn library. This plot displays the probability density of the different dose levels, allowing us to visualize the range of potential radiation doses and estimate quantiles for an individual’s exposure.

The Monte Carlo simulation method presented in this appendix provides a practical approach to estimating the uncertainty in radiation dose calculations. By generating a distribution of potential doses based on the variability of input parameters, we can derive quantiles for an individual's radiation exposure. This information can be invaluable for making informed decisions about risk management and mitigation strategies in the context of radiation protection.

A.1 Python code

```python
# Import necessary libraries and modules
import numpy as np  # Import NumPy for numerical computations
import matplotlib.pyplot as plt  # Import Matplotlib for creating plots
import seaborn as sns  # Import Seaborn for advanced data visualization

# Define a custom loguniform function
def loguniform(low=0, high=1, size=None):
    # Custom loguniform function with optional parameters
    return np.power(10, np.random.uniform(low, high, size))  # Generate random samples from loguniform distribution

# Define a custom logtriangular function
def logtriangular(low=0, central=0.5, high=1, size=None):
    # Custom logtriangular function with optional parameters
```

76
Generate random samples from logtriangular distribution

# Define probability distributions and constants
# C: radionuclide concentrations (Bq/kg)
# I: daily intake of fish (kg/day)
# DCC: dose conversion coefficient (Sv per Bq)
C = loguniform(low=3.9E-2, high=8.0E-1, size=100000) # Generate random samples of radionuclide concentrations
I = loguniform(low=5.6E-3, high=9.2E-2, size=100000) # Generate random samples of daily fish intake
EF = 3.5E+2 # Exposure Frequency (Days/year)
ED = 3.0E+1 # Exposure Duration (Years)
DCC = logtriangular(7.0E-9, 1.4E-8, 2.8E-8, 100000) # Generate random samples of dose conversion coefficients

# Determine the product of C, I, EF, ED, and DCC
# Calculate the product of the distributions and the constants EF and ED
distributions = np.array([C, I, DCC]).T # Transpose the distributions to be used in the product calculation
products = np.product(distributions, axis=1) # Compute the product of the distributions
products = products * EF * ED # Multiply the product by the exposure frequency and exposure duration

# Plot a histogram of the result using seaborn's kernel density estimation (KDE) plot
sns.kdeplot(products, color='b') # Create a KDE plot of the product distribution
plt.title('Monte Carlo Dose Distribution n=100000') # Set the title of the plot
plt.xlabel('Dose (v)') # Set the x-axis label
plt.ylabel('Probability') # Set the y-axis label
plt.show() # Display the plot
 Abbreviations and Acronyms

1981 2DMC Two-dimensional Monte Carlo
1983 BMA Bayesian Model Averaging
1984 CI Confidence interval
1985 CIM Corrected information matrix
1986 DDREF Dose and dose-rate effectiveness factor
1987 DOE U.S. Department of Energy
1988 EM expectation-maximization
1989 ERR excess relative risk
1990 GI gastrointestinal tract
1991 ICP-MS Inductively coupled plasma mass spectrometry
1992 ICRP International Commission on Radiological Protection
1993 JAEA Japanese Atomic Energy Agency
1994 JEM job-exposure matrix
1995 KDE Kernel Density Estimation
1996 LET linear energy transfer
1997 LHS Latin hypercube sampling
1999 LSS Life Span Study
2000 MAI Moment-adjusted imputation
2001 MCDS Monte Carlo Dosimetry System
2002 MCMC Markov chain Monte Carlo
2003 MCML Monte Carlo maximum likelihood
2004 MDA minimum detectable activity
2005 MIMIC Multiple Indicators, Multiple Causes
2006 MR Moment reconstruction
2007 MWC Mayak worker cohort
2008 MWDS Mayak Worker Dosimetry System
2009 NCI National Cancer Institute
2010 NTS Nevada Test Site
2011 PDF probability density function
<table>
<thead>
<tr>
<th>Year</th>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>RERF</td>
<td>Radiation Effects Research Foundation</td>
</tr>
<tr>
<td>2013</td>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>2014</td>
<td>SIMEX</td>
<td>Simulation extrapolation</td>
</tr>
<tr>
<td>2015</td>
<td>SRS</td>
<td>Simple random sampling</td>
</tr>
<tr>
<td>2016</td>
<td>STRATOS</td>
<td>STRengthening Analytical Thinking for Observational Studies</td>
</tr>
<tr>
<td>2017</td>
<td>UNSCEAR</td>
<td>United Nations Scientific Committee on the Effects of Atomic Radiation</td>
</tr>
<tr>
<td>2018</td>
<td>USRT</td>
<td>US radiologic technologists</td>
</tr>
</tbody>
</table>
Glossary

Absorbed dose ($D$): The mean energy $\delta \varepsilon$ imparted to matter of mass $dm$ by ionizing radiation at the point of interest. In the Système Internationa`l d'Unités (SI), the unit of absorbed dose is J kg$^{-1}$ with the special name gray (Gy).

Angular response: The variation of a dosimeter’s response (measured dose per unit air kerma dose) with angle of incidence of the incoming radiation track.

Bayesian methods: a set of statistical methods in which there is assumed to be some prior probability (degree of belief) distribution of some parameters of interest, and a model specified giving the probability of outcome given these parameters, these quantities being used in conjunction with Bayes theorem (Bayes and Price 1763) to derive the posterior probability distribution (degree of belief) of the parameters.

Bayesian model averaging (BMA): a specific type of Bayesian model in which a number of different models are assumed, one of which is assumed to be the true one, and inference is made on the likelihood of each model being true.

Berkson error: a type of dose measurement error in which the error in dose (=nominal dose – true dose) is assumed to be independent of the nominal dose.

Bias: if the expectation (i.e. long run average under assumed resampling in a frequentist framework) of a parameter is equal to its true value than it is said to be unbiased, otherwise it is deemed to be biased.

Bias factor: an area- and time period-specific ratio of reported dose to true air kerma that combines the calibration factor and the dosimetric factor.

Calibration factor: an average ratio of measured dose to true air kerma that characterizes the calibration practices that were used for the dosimeters that were employed in a study.

Classical error: a type of dose measurement error in which the error in dose (= nominal dose – true dose) is assumed to be independent of the true dose.

Confidence interval: in frequentist statistics, an interval of numbers for a (fixed, unknown) parameter of interest with a specified level of confidence (typically 95 %), which has the property that under theoretical repeated sampling and analysis of the sampled data, 95 % of similarly constructed confidence intervals are expected to contain the true value of the parameter.

Differential error: error that depends on the outcome of the risk regression, e.g., cancer incidence or mortality.

Dose and dose-rate effectiveness factor (DDREF): A judged factor by which the radiation effect, per unit of dose, caused by a given high or moderate dose of radiation received at high dose rates is modified.
when doses are low or are received at low dose rates. A DDREF of 2 recommended by ICRP applies to absorbed doses below 0.2 Gy and from higher absorbed doses when the dose rate is less than 0.1 Gy per hour. For more details, see ICRP Publication 103 (ICRP 2007).

**Dosimetric factor:** a ratio of reported dose to true air kerma that combines the effects of the radiation types, energies, and angles of incidence of the radiations to which the study subjects were exposed.

**Dosimetric uncertainty:** the variation in the ratio of a reported dose to the true dose.

**Energy response:** the measured dose per unit air kerma as a function of incident radiation energy, i.e., the energy dependence of measured dose due to variant absorption properties of materials in the dosimeter at different radiation energies and the intrinsic energy dependence associated with the radiation interactions producing the signal in the dosimeter.

**Excess relative risk (ERR):** Excess relative risk = risk (dose=X) / baseline risk (dose X=−) - 1.

**Functional model:** one in which some variable of interest is assumed fixed but unknown (see structural model).

**Generalized linear model (GLM):** In statistics, a generalized linear model is a flexible generalization of ordinary linear regression. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.

**Instrumental variable:** a variable that is correlated with the true dose X but is uncorrelated with the measurement error and uncorrelated with $Y - E(Y|Z,X)$.

**Job-exposure matrix (JEM):** a method of assigning dose, based on the individual being assigned a measure of dose as a result of membership of some group e.g. work location in some radiation workforce.

**Latin hypercube sampling:** a method of random sampling of $n$ variables $x_1$, ..., $x_n$ in which the probability distribution for each variable is divided into $M$ sets each of equal probability (of probability $1/M$), and $M$ samples from the $n$-dimensional hypercube $M(1) \times M(2) \times \ldots \times M(n)$ are done in such a way that there is only a single point in each of the $M$ equi-probability sets for each of the $n$ axis projections.

**Link function:** Generalized linear models include a link function that relates the expected value of the response to the linear predictors in the model. A link function transforms the probabilities of the levels of a categorical response variable to a continuous scale that is unbounded.

**Log-normal distribution:** a random variable whose logarithm is normally distributed.

**Moment reconstruction (MR):** a method to correct for error in dose by adjusting the observed exposure so that it has the same first two moments as the true exposure.
Monte Carlo: a procedure using pseudorandom numbers as a way of describing the distribution of certain quantities, dosimetric and otherwise.

Monte Carlo maximum likelihood: a method of numerically integrating the full likelihood (a function of unknown true doses, known nominal doses and unknown model parameters) over the true dose distribution to approximate the profile likelihood.

Nondifferential errors: if the conditional expectation of the outcome variable given the true and nominal dose (and some other parameters) is the same as the conditional expectation of the outcome variable given the true dose (and some other parameters) then errors are deemed nondifferential. This amounts to saying that the outcome variable only depends on the true dose – the nominal dose adds no further information.

Normal distribution: a random variable which is given by a linear scaling of the standard normal random variable, with pdf $\frac{1}{\sqrt{2\pi}} \exp \left( -\frac{x^2}{2} \right)$.

Regression calibration: a method of correction for dose error in which the expectation of the true dose given the nominal dose is substituted for the true dose in regression models.

Reliability ratio: in a simple linear model, if the independent variable $X$ has overall variance $\sigma_x^2$ with classical dose error with variance $\sigma_u^2$, the reliability ratio is $\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}$, measures the degree of attenuation of the regression coefficient towards 0. The nominal regression coefficient for $Y$ in relation to $X$ should therefore be multiplied by the inverse of the reliability ratio to correct for dose error.

Simulation extrapolation: a correction method for dose error reliant on adding a variety of defined extra amounts of classical error to the existing nominal dose estimates, and extrapolating back using the resulting estimates of some parameter of interest (e.g. ERR) to zero dose error.

Shared error: the situation in which some component of the dose for a group of persons is determined by a group of parameters in the dosimetry system, resulting in between person correlations in dose.

Structural model: one in which some unknown variable of interest can be assumed to be random (see functional model).

Taylor expansion: In mathematics, the Taylor series or Taylor expansion of an infinitely differentiable function is an infinite sum of terms that are expressed in terms of the function’s derivatives at a single point. For most common functions, the function and the sum of its Taylor series are equal near this point. The partial sum formed by the first $n + 1$ terms of a Taylor series is a polynomial of degree $n$ that is called the $n$th Taylor polynomial of the function. Taylor polynomials are approximations of a function, which become generally more accurate as $n$ increases. Taylor’s theorem gives quantitative estimates on the error introduced by the use of such approximations.
Unshared error: the situation in which some component of dose is determined solely by individual variations in certain underlying parameters, so that there are no correlations in this component of dose between individuals.
References


Thierry-Cheff I, Richardson DB, Daniels RD, Gillies M, Hamra GB, Haylock R, Kesminiene A, Laurier D, Leuraud K, Moissonnier M et al. 2015. Dose Estimation for a Study of Nuclear Workers in France, the


