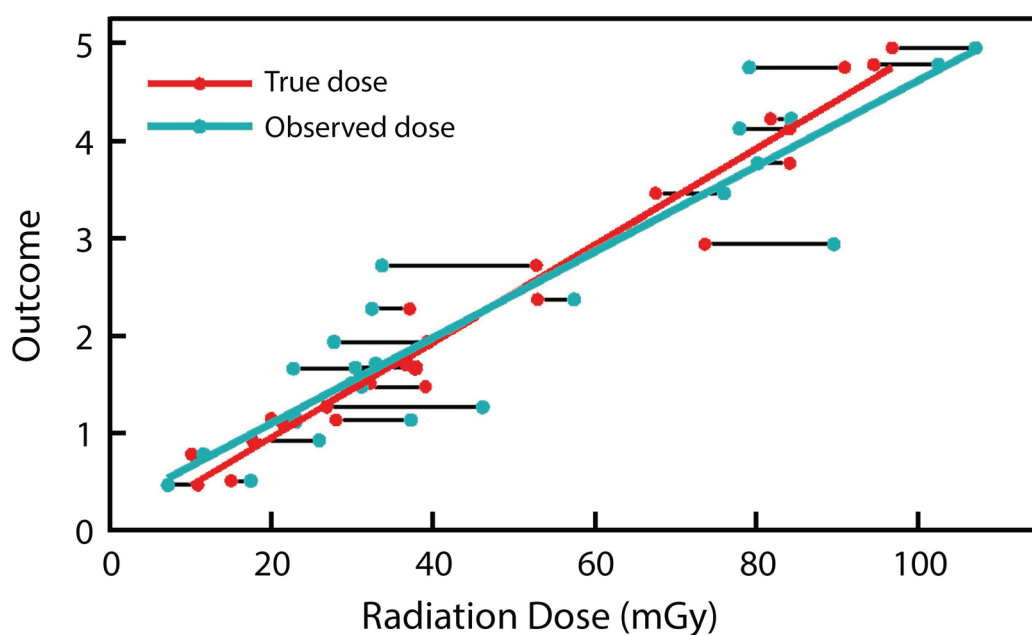


RECOMMENDATIONS ON STATISTICAL APPROACHES TO ACCOUNT FOR DOSE UNCERTAINTIES IN RADIATION EPIDEMIOLOGIC RISK MODELS



National Council on Radiation Protection and Measurements



Documents

Document Sharing Hub

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

May 10, 2024

**National Council on Radiation Protection and Measurements
7910 Woodmont Avenue, Suite 905 / Bethesda, Maryland 20814-3046**

LEGAL NOTICE

This Commentary was prepared by the National Council on Radiation Protection and Measurements (NCRP). The Council strives to provide accurate, complete and useful information in its documents. However, neither NCRP, the members of NCRP, other persons contributing to or assisting in the preparation of this Commentary, nor any person acting on the behalf of any of these parties: (a) makes any warranty or representation, express or implied, with respect to the accuracy, completeness or usefulness of the information contained in this Commentary, or that the use of any information, method or process disclosed in this Commentary may not infringe on privately owned rights; or (b) assumes any liability with respect to the use of, or for damages resulting from the use of any information, method or process disclosed in this Commentary, *under the Civil Rights Act of 1964, Section 701 et seq. as amended 42 U.S.C. Section 2000e et seq. (Title VII) or any other statutory or common law theory governing liability.*

Disclaimer

Any mention of commercial products within NCRP publications is for information only; it does not imply recommendation or endorsement by NCRP. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the NCRP and do not necessarily reflect the views of any of the authors' organizations or agencies.

Library of Congress Cataloging-in-Publication Data

Names: National Council on Radiation Protection and Measurements publisher.

Title: Recommendations on statistical approaches to account for dose uncertainties in radiation epidemiologic risk models : recommendations of the National Council on Radiation Protection and Measurements / National Council on Radiation Protection and Measurements. Other titles: NCRP commentary ; no. 34.

Description: Bethesda, Maryland : National Council on Radiation Protection and Measurements, May 1, 2024. | Series: NCRP commentary ; no. 34 | Includes bibliographical references. | Summary: "Accurate exposure estimation in radiation epidemiologic studies is essential for reliable health risk assessment. Failure to account appropriately for uncertainties in dose estimation and model assumptions could lead to biased results in the evaluation of the radiation dose-response as well as incorrect confidence bounds for risk parameters. 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 estimation and inference for the dose-response. The purpose of this Commentary is to provide a guide regarding 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 disadvantages of the methods. This Commentary addresses studies of external and internal exposures and provides guidance on both shared and unshared uncertainty in the estimation of absorbed dose. Of particular interest are statistical methods for assessing dose-response in epidemiologic studies of internal emitters, for which doses are calculated using exposure and retention models with many parameters. Each parameter is associated with various sources and amounts of uncertainty" -- Provided by publisher.

Identifiers: LCCN 2024013853 (print) | LCCN 2024013854 (ebook) | ISBN 9781944888367 (hardcover) | ISBN 9781944888374 (PDF)

Subjects: MESH: Dose-Response Relationship, Radiation | Radiometry--methods | Models, Statistical | Epidemiologic Studies | United States

Classification: LCC R906 (print) | LCC R906 (ebook) | NLM WN 665 | DDC 612/.014480287--dc23/eng/20240410

LC record available at <https://lcn.loc.gov/2024013853>

LC ebook record available at <https://lcn.loc.gov/2024013854>

Copyright © National Council on Radiation Protection and Measurements 2024

All rights reserved. This publication is protected by copyright. No part of this publication may be reproduced in any form or by any means, including photocopying, or utilized by any information storage and retrieval system without written permission from the copyright owner, except for brief quotation in critical articles or reviews.

[For detailed information on the availability of this and other NCRP publications see page 83.]

Preface

Accurate exposure estimation in radiation epidemiologic studies is essential for reliable health risk assessment. Failure to account appropriately for uncertainties in dose estimation and model assumptions could lead to biased results in the evaluation of the radiation dose-response as well as incorrect confidence bounds for risk parameters. 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 estimation and inference for the dose-response. The purpose of this Commentary is to provide a guide regarding 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 disadvantages of the methods. This Commentary addresses studies of external and internal exposures and provides guidance on both shared and unshared uncertainty in the estimation of absorbed dose. Of particular interest are statistical methods for assessing dose-response in epidemiologic studies of internal emitters, for which doses are calculated using exposure and retention models with many parameters. Each parameter is associated with various sources and amounts of uncertainty.

This Commentary draws from and builds upon previous National Council on Radiation Protection and Measurements (NCRP) commentaries and reports on closely related topics, including:

- Commentary No. 14, *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination* (1996)
- Report No. 158, *Uncertainties in the Measurement and Dosimetry of External Radiation* (2007)
- Report No. 164, *Uncertainties in Internal Radiation Dose Assessment* (2009)
- Report No. 171, *Uncertainties in the Estimation of Radiation Risks and Probability of Disease Causation* (2012)
- Report No. 178, *Deriving Organ Doses and Their Uncertainty for Epidemiologic Studies* (2018)

The intended audience for this Commentary is dosimetrists, statisticians, epidemiologists, organizations, and decision makers 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:

Co-Chairs

Jonine L. Bernstein
Memorial Sloan-Kettering Cancer Center
New York, New York

Harry M. Cullings
Radiation Research Effects Foundation
Hiroshima, Japan

Members

Michael B. Bellamy

Memorial Sloan-Kettering Cancer Center
New York, New York

Mark P. Little

National Cancer Institute
Bethesda, Maryland

Benjamin French

Vanderbilt University Medical Center
Nashville, Tennessee

Carmen D. Tekwe

Indiana University School of Public
Health
Bloomington, Indiana

NCRP Secretariat

Helen A. Grogan, *Staff Consultant*

Laura J. Atwell, *Director of Operations*

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. The Council expresses its gratitude to the presenters and participants at the Workshop (Appendix A) hosted by NCRP at the start of the project and to the subject matter experts who provided valuable feedback on a draft of the Commentary.

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.

Kathryn D. Held
President 2019 – 2024

Kathryn A. Higley
President

Contents

Preface	iii
1. Executive Summary	1
1.1 Background	1
1.2 Characteristics of Dosimetric Uncertainty	1
1.3 Effects of Dosimetric Uncertainty on Radiation Epidemiologic Studies	1
1.4 Statistical Methods to Account for Dosimetric Uncertainty	2
1.5 Conclusions and Recommendations	3
2. Introduction	4
3. Definition of Variability, Uncertainty, and Error Types	5
3.1 Shared and Unshared Error	6
3.2 Classical or Berkson Error	7
3.3 Other Types of Uncertainty	8
4. Dosimetric Uncertainties in Radiation Epidemiologic Studies	10
4.1 Sources of Uncertainty in External Dosimetry	11
4.1.1 Japanese Atomic-Bomb Survivors	13
4.1.2 IARC 15-Country Study and International Nuclear Worker Study	15
4.2 Sources of Uncertainty in Internal Dosimetry	15
4.2.1 Mayak Nuclear Workers	18
4.2.2 Wismut Cohort of Uranium Miners	18
4.3 General Approach to Uncertainty Analysis Based on Dosimetric Parameters .	19
4.4 Numerical Models for Error Propagation in Uncertainty Analysis	21
4.5 Monte Carlo Analysis	22
4.5.1 Example 1: Cohorts with Negligible Shared Error	22
4.5.2 Example 2: Cohorts with Non-Negligible Shared Error	23
5. Effects of Dose Uncertainties in Radiation Epidemiologic Risk Models	27
6. Methods to Account for Dose Uncertainty in Radiation Risk Models	29
6.1 Simulation Extrapolation	29
6.1.1 Type of Dose Data to Which SIMEX Can be Applied	29
6.1.2 How SIMEX Works	30
6.1.3 Assumptions	30
6.1.4 Procedure for Applying SIMEX	30
6.1.5 Strengths and Limitations of SIMEX	31
6.2 Regression Calibration	31
6.2.1 Type of Dose Data to Which Regression Calibration Can be Applied ..	32
6.2.2 How Regression Calibration Works	32
6.2.3 Assumptions	32
6.2.4 Procedure for Applying Regression Calibration	33
6.2.5 Strengths and Limitations of Regression Calibration	33

6.3	Likelihood-Based Methods and Monte Carlo Maximum Likelihood	34
6.3.1	Background on Use of Likelihood in Regression	34
6.3.2	Type of Dose Data to Which Likelihood-Based Methods and MCML Can be Applied	35
6.3.3	How Likelihood-Based Methods and MCML Work	35
6.3.4	Assumptions	36
6.3.5	Procedure for Applying Likelihood-Based Methods and MCML	36
6.3.6	Strengths and Limitations of Likelihood-Based Methods and MCML	36
6.4	Bayesian Markov Chain Monte Carlo	36
6.4.1	Type of Dose Data to Which Bayesian MCMC Can be Applied	37
6.4.2	How Bayesian MCMC Works	37
6.4.3	Assumptions	37
6.4.4	Procedure for Applying Bayesian MCMC	37
6.4.5	Strengths and Limitations of Bayesian MCMC	37
6.5	Two-Dimensional Monte Carlo with Bayesian Model Averaging	38
6.5.1	Type of Dose Data to Which 2DMC With BMA Can be Applied	38
6.5.2	How 2DMC With BMA Works	38
6.5.3	Assumptions	39
6.5.4	Procedure for Applying 2DMC With BMA	39
6.5.5	Strengths and Limitations of 2DMC With BMA	39
6.6	Frequentist Model Averaging	40
6.6.1	Type of Dose Data to Which FMA Can be Applied	40
6.6.2	How FMA Works	40
6.6.3	Assumptions	41
6.6.4	Procedure for Applying FMA	41
6.6.5	Strengths and Limitations of FMA	41
6.7	Corrected Information Matrix	42
6.7.1	Type of Dose Data to Which CIM Can be Applied	42
6.7.2	How CIM Works	42
6.7.3	Assumptions	42
6.7.4	Procedure for Applying CIM	42
6.7.5	Strengths and Limitations of CIM	43
6.8	Moment Reconstruction and Moment-Adjusted Imputation	43
6.8.1	Type of Dose Data to Which MR and MAI Can be Applied	43
6.8.2	How MR and MAI Work	43
6.8.3	Assumptions	44
6.8.4	Procedure for Applying MR and MAI	44
6.8.5	Strengths and Limitations of MR and MAI	44
6.9	Summary	44
7.	Summary and Recommendations	49
	Appendix A. Scientific Committee 1-28 Workshop	52
	Appendix B. A Sample Monte Carlo Simulation to Determine Uncertainty in Radiation Dose Under Conditions of Negligible Shared Error	54
B.1	Libraries and Modules	54
B.2	Custom Density Functions	55
B.3	Probability Density Functions and Constants	55
B.4	Monte Carlo Simulation	55

B.5 Visualization 55

B.6 Python Code 56

Abbreviations, Acronyms and Symbols 57

Glossary 58

References 61

Scientific Committee 71

The NCRP 74

NCRP Publications 83

1. Executive Summary

1.1 Background

Absorbed dose from exposure of the human body to ionizing radiation cannot be measured directly, and doses to specific organs must be estimated based on indirect measurements and mathematical or biophysical modeling. Thus, uncertainty in dosimetry is inherently part of radiation health effects studies. Failure to account for dosimetric uncertainties has the potential to distort relationships between radiation dose and outcomes in dose-response models. The purpose of this Commentary is to evaluate multiple, commonly used, statistical approaches used to account for uncertainties in dosimetry estimates and review the application of various uncertainty estimation techniques. Previous National Council on Radiation Protection and Measurements (NCRP) reports have examined and characterized dosimetric uncertainty in external and internal exposures to ionizing radiation (NCRP 2007, 2009a), and uncertainties in study results. No NCRP document has thoroughly reviewed the application of multiple statistical methods to adjust for dosimetric uncertainty. This Commentary aims to thoroughly investigate a variety of prominent methods used to account for dose uncertainty.

1.2 Characteristics of Dosimetric Uncertainty

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. Random uncertainties can either be of classical (for which the measured dose is assumed to be derived from the true dose by adding an error which is independent of the true dose) or of Berkson type (for which the true dose is assumed to be derived from the measured dose by adding an error term which is independent of the measured dose), as determined by the statistical relationship between true dose and the estimated dose. 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. Some studies are based on dosimetry with complex uncertainty, comprising a mixture of Berkson and classical error. Complex uncertainty is associated with many studies, particularly studies 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, 2009a).

1.3 Effects of Dosimetric Uncertainty on Radiation Epidemiologic 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 nondifferential 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; Keogh et al. 2020; van Smeden et al. 2020). It is these latter effects on the slope or curvature of the dose response for which statistical adjustments can be made. The statistical methods for accounting for radiation dose uncertainty in dose-response curves are reviewed here.

1.4 Statistical Methods to Account for Dosimetric Uncertainty

This Commentary reviews the strengths and weaknesses of eight different commonly used statistical methods for characterizing uncertainty prominent in the current literature:

- simulation extrapolation;
- regression calibration;
- likelihood-based methods;
- Markov chain Monte Carlo;
- two-dimensional Monte Carlo with Bayesian model averaging;
- frequentist model averaging;
- corrected information matrix; and
- moment reconstruction and moment-adjusted imputation.

Simulation extrapolation (SIMEX) is a conceptually straightforward method in which increasing amounts of randomly generated uncertainty are added to dose estimates and a risk regression is repeatedly refitted, resulting in a relationship that is extrapolated to zero dose uncertainty. *Regression calibration* is a popular analytical method in which the risk is assessed by replacing the true dose in regression models by the expected value of the (unknown) 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 between true and uncertain dose and between true dose and epidemiologic outcome. In many current studies, doses are often estimated by Monte Carlo methods making use of the complex dosimetric uncertainty models that have been developed. *Markov chain Monte Carlo* (MCMC) is a Bayesian likelihood-based method that assumes prior distributions for parameters including radiation risk and creates a posterior distribution by successive applications of Bayes' theorem. MCMC is solved by Monte Carlo methods, as a closed-form solution is generally analytically intractable. *Two-dimensional Monte Carlo* (2DMC) 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 data set, then combines the results in an average weighted by their relative goodness of fit — hence the BMA. *Frequentist model averaging* (FMA) shares some features with 2DMC with BMA, and is designed for handling the same type of data, with a mixture of shared and unshared errors, but uses a fundamentally frequentist framework that averages randomly generated results in numbers proportional to their relative goodness of fit. *Corrected information matrix* (CIM) is a method that uses the score statistic to adjust the confidence intervals of the risk estimates for 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

Uncertainties in dose are ubiquitous in radiation epidemiology studies, and failure to adjust for dose uncertainties in statistical analyses can result in incorrect estimation and inference regarding radiation health effects. Each analysis method to adjust for dosimetric uncertainty has advantages and disadvantages that need to be considered in each individual application. Below is a summary of the conclusions and associated recommendations in this Commentary.

- Whenever possible, a rigorous analysis of dosimetric uncertainty should be included in every study, and statistical methods should be used to account for dosimetric uncertainty when estimating the radiation dose-response.
- Regression calibration, MR, MCML, Bayesian MCMC and FMA are the most generally applicable methods used to account for dosimetric uncertainty. MR and MAI are most suited to problems in which there is thought to be a nonlinear dose response, or in which errors are nondifferential (not dependent on the relationship between true dose and outcome).
- CIM and SIMEX are limited in the problems to which they can be applied (*i.e.*, models with only Berkson uncertainty in the first case and only classical uncertainty in the second). In both cases the magnitude of errors must be known, and for the CIM method the error correlation structure must also be known. Furthermore, SIMEX is computationally somewhat cumbersome.
- Methods such as MCML and Bayesian MCMC are computationally intensive and may be prohibitively demanding of computer resources for large cohorts and complex models of dosimetric uncertainty.
- 2DMC with BMA is quite general but has decisive statistical problems that need to be considered.
- A software repository should be created for as many methods of analysis as possible. Such methods and computational programs should be accessible to the public.
- More research, and working groups, should be devoted to developing and evaluating methods to account for dosimetric uncertainty in combination with model uncertainty and outcome misclassification.

2. Introduction

There is direct evidence of human health risks following moderate and high levels of ionizing radiation absorbed dose (NA/NRC 2006; UNSCEAR 2008, 2015; IARC 2012; NCRP 2012, 2018a). Several studies of highly radiogenic cancers such as leukemia and thyroid cancer (notably following childhood exposures) (Lubin et al. 2017; Little et al. 2018; Little, Wakeford, Bouffler et al. 2022a, 2022b) report estimates of radiation risk at low doses (*i.e.*, <100 mGy). Groups exposed to moderate and high levels of absorbed radiation are utilized to extrapolate cancer risks and outcomes other than cancer from low doses of radiation. These studies have been reviewed by NCRP (2018a) and in a monograph published by the National Cancer Institute (Berrington de Gonzalez et al. 2020; Daniels et al. 2020; Gilbert et al. 2020; Hauptmann et al. 2020; Linet et al. 2020; Schubauer-Berigan 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 risks at low doses and low dose rates. Statistical modeling is crucial to the resolution of the uncertainties in the dose-response relationship and accounting for both systematic and random dosimetric errors in dose-response models. Assessment of dosimetric uncertainty is vital for deriving unbiased risks from Japanese atomic-bomb survivor studies often used to quantify population risks (ICRP 2007; UNSCEAR 2008; Ozasa et al. 2019).

Dosimetric uncertainty arises from uncertainty in the parameters and structure of the dosimetry system employed to estimate radiation dose and from measurement error in data input to the dosimetry system. Quantifying dosimetric uncertainty when estimating dose-response relationships has been the subject of much interest in biostatistics (Pierce et al. 1990, 1992). Dosimetric uncertainty can alter the dose response relationship substantially and hence the derived population risk estimates (Carroll et al. 2006).

This Commentary focuses on statistical methods to account for dosimetric uncertainty in regression models. It does 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 (UNSCEAR 2018). NCRP Report No. 171 (NCRP 2012) delineates these other important random and systemic errors including methodological and epidemiological uncertainties (*i.e.*, biases and confounding influences inherent in study design that are not often quantifiable, uncertainties in lifetime risk projections, and overall uncertainties in risk estimates that are used for radiation protection purposes). This Commentary surveys assessments of quantification of uncertainties in dosimetry and considers the effects of dosimetric uncertainties in risk regression models, and statistical methods to account for these uncertainties. The dosimetric uncertainty issues addressed in this Commentary are found in radiation epidemiologic studies of exposure from many sources including occupational, environmental, medical exposure, and space radiation. Throughout the Commentary, examples from large and well-characterized epidemiologic cohorts are provided 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 nonionizing radiation. The Commentary ends with conclusions on recommended methods to account for dosimetric uncertainty.

3. Definition of Variability, Uncertainty, and Error Types

Variability refers most generally to any variation or nonuniformity among values of some quantity that is assumed to apply to members of a cohort. Uncertainty refers to a lack of knowledge of some or all of the variability (*i.e.*, not knowing the exact values of some quantities). Error refers to any difference between true and observed or assumed values of some quantity, and random error, which corresponds to uncertainty, is an error whose size is unknown and is assumed to follow a probability distribution. A random error may have a non-zero mean, in which case it confers a bias and is referred to herein as a systematic error. Thus, variability, which can be observed, consists of variation in some true quantity plus error, which cannot be observed separately (NCRP 1996). Furthermore, random and systematic errors can be classified as either shared or unshared, and each of these may also be classified as classical or Berkson form (NCRP 2007, 2009a). These are illustrated in Figure 3.1.

Variability 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 family members may vary with age, sex, and socioeconomic status. Common measures of variability include range, variance, and standard deviation (SD), which provide insight into the spread or dispersion of the data. Uncertainty, on the other hand, can arise due to 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

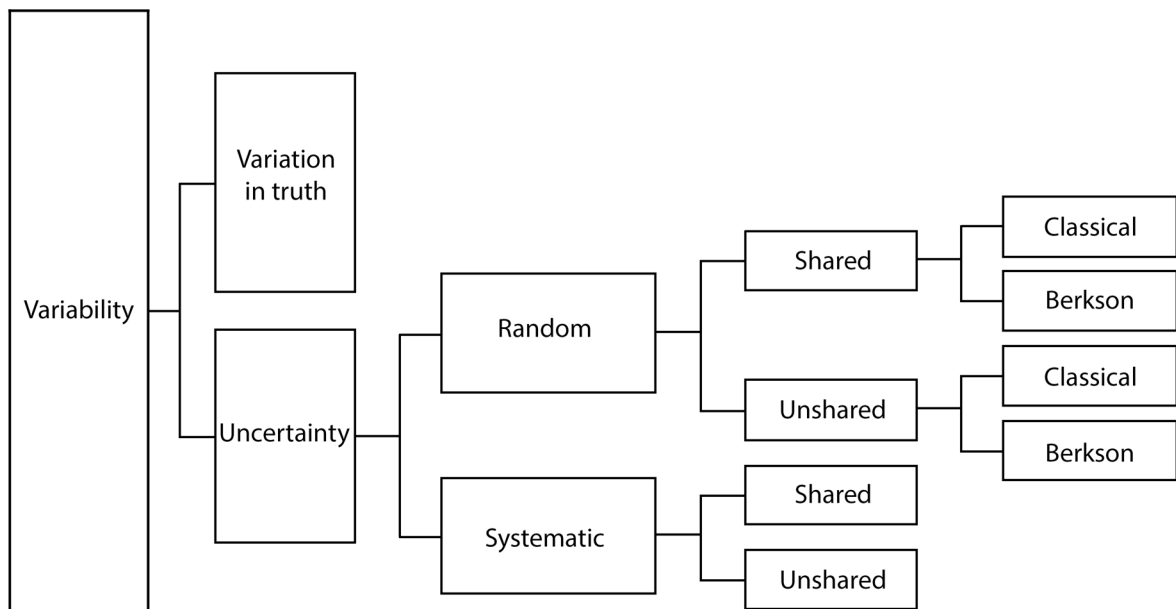


Fig 3.1. The relationship among different types of uncertainty in radiation dosimetry.

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.

Uncertainty can be either *random* or *systematic*. Random uncertainty typically arises from statistical fluctuations that are inherent to any measurement of a particular quantity. That is, repeated measurements under identical conditions produce a distribution of results that represent the operation of unknown processes that can be thought of as fundamentally probabilistic or random in nature. Systematic uncertainty 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 non-zero (but possibly unknown) value, which will generally result in bias. The term “bias” can apply to any kind of random error, shared or unshared, classical or Berkson. Even an error that is shared equally among all members of a cohort could be modeled as having a distribution with non-zero mean, and therefore a bias, if for example values were being generated for a cohort by Monte Carlo methods. On the other hand, if a single realization of a shared error is non-zero, herein this is not referred to as a bias, but only as a shared error. The term “bias” is reserved for a situation in which some quantity is *measured with error, the mean of which is non-zero*. Although shared errors may produce patterns in the error structure for a cohort and may be thought of as being “systematic” in that sense, such errors are not referred to as systematic errors in this document unless they have non-zero mean.

3.1 Shared and Unshared Error

In dose reconstruction scenarios it is 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 that is 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 (*i.e.*, perfect). 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 the weapon explosion. However, members of these crews do not necessarily share uncertainties related to their individual positions on the ship. Similarly, a ground unit present during a nuclear weapons test at the Nevada Test Site share uncertainties in weapon yield and position, 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 following the test. In occupational exposure scenarios, available data typically include dates of employment, job title, and duration of employment, as well as dosimeter readings and bioassay measurements. However, workers at the same site, and in the same building, with the same job title (*e.g.*, chemical operator), may have 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 results will be shared if the same 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.

3.2 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:

$$\text{estimate} = \text{true value} + \text{measurement uncertainty}, \quad (3.1)$$

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 conducted 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).

The Berkson error model, in which the true value varies from the estimate by an uncertainty that is random and is independent of the estimate, can be written:

$$\text{true value} = \text{estimate} + \text{individual deviation}, \quad (3.2)$$

where: *individual deviation* is a random variable (with mean zero) and is independent of *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 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 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 assumed to be well known or known with small uncertainty.

3.3 Other Types of Uncertainty

This Commentary primarily 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. NCRP Report No. 171 (NCRP 2012) provides a fuller discussion of nondosimetric uncertainties such as flaws in study design.

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. 1998). It is possible to account for uncertainties in the various types of models employed. 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 BMA 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 criticized, and MMI in particular may not adequately take account of full uncertainty that a more fully Bayesian method may supply resulting from uncertainties in the underlying models (Wang et al. 2012). For example, Little and Wakeford (2013) pointed out that MMI 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 Life Span Study (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) (UNSCEAR 2018). 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. For example, Sposto et al. (1992) used expectation-maximization 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 adjust for outcome misclassification. However, even accounting for death certificate causes of death based on autopsy evaluations, spurious associations might arise if autopsy rates are correlated with dose (NCRP 2012).

4. Dosimetric Uncertainties in Radiation Epidemiologic Studies

In many assessments, the estimated quantity can be a discrepant estimate of the true value, which can itself be subject to 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 inaccuracy 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 the radionuclide was actually in an insoluble form.

Measurement errors in exposure assessment, and resultant dosimetric uncertainties are inherent in many radiation epidemiologic studies (NCRP 1996, 1997; Bellamy et al. 2023). Such errors arise partially from uncertainties in parameters and model structure in complex dosimetry systems. Sources of uncertainty can be propagated through the exposure assessment to produce an estimate of uncertainty in the final dosimetry endpoint. While dosimetric uncertainty is unavoidable, it is possible to characterize sources of uncertainty in the inputs of a dosimetric evaluation. Personal dosimeters measure external radiation dose at a point outside the body, not in the tissue, and other uncertainties exist in addition to those measured by the dosimeter (NCRP 2018b). Similarly, bioassays may have considerable uncertainty in the estimates of dose to tissues that derive from estimates of radionuclide concentrations in the tissues of concern. Unfortunately, many studies do not have dose information such as personal dosimeters or measurements such as bioassay on the entire cohort.

The following section introduces 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 necessary for generating dependable risk assessments and facilitating well-informed public health decisions. It is important 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; UNSCEAR 2015).

Various methods and techniques, such as Monte Carlo simulations and expert judgment, are used to address uncertainty in radiation epidemiology research (NCRP 2018b). Monte Carlo simulations offer insight into the probability distribution of uncertainties and their possible influence on dose estimations (NCRP 1997), while expert judgment is vital to determine parameter uncertainties, particularly when data are limited or incomplete (NCRP 1996,

2009a). 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 (IAEA 1989; NCRP 1996, 2009a, 2009b).

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 when addressing 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 often derived using available evidence on the model parameters, the model structure, and any dependencies that exist among model components, supplemented in many 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. This type of uncertainty does not necessarily lead to bias, although bias can result if the mean of the associated distribution differs from the fixed value assigned to it, or if the distribution assumed differs in mean from the true underlying distribution.

When relevant cohort-specific data are unavailable, a defensible method for obtaining subjective probability distributions is through the formal elicitation of expert judgment (expert elicitation) using a structured questionnaire or interview (NCRP 1996; Little et al. 1997a, 1997b). Formal elicitation is used to encode what is known about a model component. Because of the expense involved in conducting a formal elicitation, 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 expert elicitation, important assessments should be assigned to at least two independent organizations, and the assessment should include resources to resolve discrepancies. Assessment using two independent organizations 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 2018b) 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 are discussed in these reports.

An individual can be exposed in different ways to radiation sources outside the body with a variety of ionizing radiations such as x rays, gamma rays, neutrons, protons and beta particles.

Measurements can be made of radiation fields resulting from these radiation sources. The measured quantity may be the number of ionizations in a mass or air kerma 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 measured quantity to the expected absorbed dose in tissue (D_T). Estimated absorbed dose, D_T , will always be uncertain, due to the error in the underlying measurement and uncertainty in the model and model parameters used to relate the measured quantity to the absorbed dose, D_T , in a body organ of a particular individual. The relationship between a measured or specified quantity in air and D_T can be estimated by measurements using a tissue-equivalent phantom that is a physical mockup of the human body (Stovall et al. 2006; Howell et al. 2019). 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 measuring the radiation field and 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 or gamma rays) in a room with workers. The estimated photon flux incident on the workers is used to estimate relevant organ doses. Uncertainties are associated with characterization of the radiation field and the subsequent estimation of organ doses.

Radiation field measurements can be conducted using either area monitors or personal monitors. The major sources of uncertainty arise from imperfect knowledge of the radiation field for which the instrument was employed. Improper assumptions concerning the radiation field may lead to differences in energy response and angular response that are assumed for the area monitors or personal dosimeters (ICRP 2010). Other significant sources of measurement error may arise from readout bias, limited resolution, improper calibration procedures 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.

Although a radiation field may be 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 radiation field assumptions. Dosimetry models have parameters that cannot be known with certainty. For example, the 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 is important for dose modeling. Improperly characterizing a 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 doses are estimated from either area measurements or personal dosimeter readings.

Area (field) measurements characterize the radiation field rather than the exposure of a specific individual. Area measurements can be generalized to any individual moving through the radiation field. Area measurements are typically 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. Area measurements have the advantage of collecting types of data that would be impractical as personal measurements (e.g., collection of gamma-spectral data by *in situ* gamma spectrometry). Area measurement detectors also have less restrictions on size and weight; therefore they can be manufactured with larger, more sensitive active volumes and generally provide more precise, accurate and multimodal results relative to the smaller, lighter personal detectors. Area measurements can also provide data over periods of time

much longer than would be practical for personal measurements and, hence, increase precision of the measurements.

Personal-dosimeter measurements more closely represent the dose to an individual than area measurements. Personal dosimeters are typically passive devices worn on the body which provide dose information after processing. Instantaneous and integrating solid state electronic dosimeters are also available, which provide information on 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 with longer counting (observation) times.

Personal measurements are less useful to generalize to others than the persons wearing them, except possibly for cases where other persons moved in the field through the same locations for the same time periods. Personal measurements as a surrogate for group exposure may be an acceptable assumption only for certain situations (*e.g.*, estimating the average dose for military troops traveling on foot). Personal measurements were generalized for military personnel participating in early nuclear-weapons tests. Military individuals were usually not issued individual dosimeters during maneuvers at the test sites, but rather their radiation exposure was inferred from a limited number of cohort dosimeters issued to a few members of their unit (NA/NRC 1989, 2003). Although not every member of military units traversed the exact same path through a radiation field, the differences in dose and the related uncertainty for individuals taking slightly different paths may or may not be important. Lack of other relevant dosimetry information such as inhalation of resuspended radioactive debris (Boice, Cohen, Mumma, Chen et al. 2022) could introduce greater uncertainty.

With the complexity of the absorbed dose calculation from personal measurements in mind, it is important to consider the various sources of uncertainty when calculating organ dose. The complexity of dose calculations from personal measurements is compounded by the fact that radiation fields are 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," estimated from personal measurements.

The following sections provide 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 International Nuclear Worker Study (INWORKS) (Thierry-Chef et al. 2015)].

4.1.1 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 DS02R1 to reflect the review and selection among multiple versions of survivor exposure records as well as revisions in terrain

shielding (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 considering their reported locations at the times of the bombings, using detailed calculations of the radiation fluences at various distances from the detonation sites along with structural shielding models of houses, and human phantoms to account 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 not been included in the dosimetry to date by RERF (Roesch 1987; Young and Kerr 2005) and many others evaluating risks in the atomic-bomb survivors, although some authors have suggested that such doses should 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, the U.S. National Cancer Institute and the Japan Atomic Energy Agency to construct a modern set of human phantoms (Griffin et al. 2019; Paulbeck et al. 2019) to replace the crude human phantoms fashioned 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 detailed analysis of uncertainties in DS02 is provided in Young and Kerr (2005), starting with uncertainties in quantities such as the estimated heights of burst of the two bombs. Many of the related errors are shared among all or subgroups of survivors. The uncertainties in survivor dosimetry are dominated by errors in survivors' reported locations and to a lesser extent their reported shielding by buildings (Jablon 1971; Cullings et al. 2017). An adjustment based on regression calibration (Section 6.2) with an assumed 35 % geometric standard deviation classical error (where $\ln[\text{dose}]$ has classical error with SD 0.35) has long been used at RERF and by many associated investigators (Pierce et al. 1990, 1992; Little and Muirhead 1996, 1997, 1998, 2000). A somewhat different adjustment based on 20 % Berkson error and 40 % classical error has been proposed (Pierce et al. 2008) but has not been adopted in other analyses because of issues regarding the size of the Berkson error and the use of one Berkson error value for all survivors (see below). An analysis of dyslipidemia in a small cohort of high-dose atomic-bomb survivors with multiple indicators, multiple causes 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 in a shielded configuration 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. Survivors 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:

- inside of houses;
- outside of houses;
- completely in the open; or
- in Nagasaki factories (Cullings et al. 2006).

Survivors without a shielding history therefore 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 >400,000 nuclear workers at 154 nuclear facilities, 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 response of dosimeters in workplace exposure conditions and historical calibration practices for photons in the range of 100 to 3,000 keV were identified as the main sources of errors. Uncertainties related to dosimetry technology and radiation fields were quantified to derive period- and facility-specific estimates of bias and uncertainties in recorded doses. Bias estimates were based on:

- evaluation of predominant workplace radiation from measurement studies and dosimetry expert assessment; and
- 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). Errors 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 postulated to be shared Berkson in form (Thierry-Chef et al. 2007). The only classical errors in the dosimetry relate to dosimeter processing and reading, which were postulated to be of unshared form, but are also regarded as minor compared to other errors, which were all of Berkson form (Thierry-Chef et al. 2007).

These dosimetric considerations and limitations to some extent carry over directly to INWORKS (Thierry-Chef et al. 2015), the successor to the 15-Country study, which assessed only nuclear workers in France, United Kingdom, and United States, 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), although use was not made of them in the final organ dose estimates.

4.2 Sources of Uncertainty in Internal Dosimetry

Each component included in the process of dose determination following intakes of radionuclides 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. NCRP Report No. 164 (NCRP 2009a) and NCRP Report No. 178 (NCRP 2018b) provide 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 measure the absorbed dose to a human organ directly, all internal doses must be estimated based on indirect measurements (e.g., bioassay data, environmental data). Energy deposition in receptor organs is estimated using mathematical models that simulate the bioaccumulation of the radionuclide and the transfer of energy between source and target regions. Internal dose determination inherently contains many uncertainties. Internal dose is based on models often built on the basis of small data sets over limited cohorts, and sometimes only partially-relevant data with varying levels of reliability (Leggett and Williams 1981). Improvements to dosimetric models may, depending on availability, include incorporation of autopsy results for occupationally or environmentally exposed subjects to assess the solubility of inhaled material as well as the relative contents in organs and tissues (Leggett et al. 2022; Boice, Cohen, Mumma, Golden et al. 2022; Samuels and Leggett 2023).

Internal dosimetry assessment uncertainties may arise in several broad categories. These include uncertainties in measurements, intakes, biokinetic model structure and parameters, dosimetric model parameters, and methods used to 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 also arise in sampling. The critical issues are 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.

Other large uncertainties, known and unknown, might arise from missing or absent monitoring as well as contaminated urine samples (NCRP 2018b). Reuse and contamination of glass collection bottles, for example, have been reported in the early operational years at a U.K. plutonium facility (Riddell et al. 2000).

Indirect bioassay measurements for activity levels approaching the minimum detectable activity, including all forms of error, are characterized by varied levels of measurement uncertainties. Several examples illustrate this point for:

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

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 to the subject being measured, a source of shared error; and
- distribution of activity in an organ being counted.

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). As noted above, 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).

Radiation doses arising from inhalation occur in occupational settings and in environmental circumstances. An important source of uncertainty in estimation of doses from inhalation is the physico-chemical form of the radionuclide. Radionuclides are usually assumed to be attached to aerosol particles. However, this is not always true, such as 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 solubility (Dorrian 1997; Malá et al. 2013). In environmental exposure situations, particle 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), 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, a source of shared error.

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 preferably from measurements of radionuclide concentrations in the consumed foodstuffs or from the use of mathematical models to estimate concentrations and uptakes and the consumption rates of those foodstuffs. For example, absorbed dose to the thyroid of children living around Hanford during releases of radioactive ^{131}I and other radionuclides, were based primarily on milk consumption rates and detailed characterization of ^{131}I concentrations in air, food and milk products, and on ground surface (Davis et al. 2004).

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:

- there is a large variety of foodstuffs in each of those categories;
- consumption rates of individual foodstuffs vary widely; and
- 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 tract uptake, and the variation between foods consumed by various demographic subgroups.

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 relies on professional judgment, a largely subjective process. Thus, it is not surprising that different analysts may produce different estimates or characterizations of dose and of uncertainty in dose for the same dose assessment endpoint. Since two analysts may produce different results, it is useful 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 (former USSR) in which nuclear weapons were assembled, including plutonium cores. Koshurnikova et al. (1999) provides background information about the Mayak Nuclear Workers. In the Mayak worker cohort, 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. 2019). Mayak worker cohort 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. 2019) 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 expressed 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 describing the behavior of plutonium in the body and dosimetry models (Birchall et al. 2017; Vostrotin et al. 2019). 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 (Vostrotin et al. 2019). It is thought that the errors in dose in the MWDS are mostly of shared Berkson form (Zhang et al. 2017; Stram et al. 2021a). 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. 2021a).

4.2.2 *Wismut Cohort of Uranium Miners*

The Wismut cohort was established in the 1990s and includes about 59,000 former male employees of the Wismut company (follow-up period 1946 to 2018) (Kreuzer et al. 2002, 2010). For each worker, a working history was derived from the payrolls of the Wismut company. In the late 1990s, the investigators developed a detailed job-exposure matrix for estimating the exposure to radon progeny for employees of Wismut. The job-exposure matrix was further developed for scientific purposes and implemented in a software program in the mid-2000s (van Dillen et al. 2011; Küchenhoff et al. 2018; Deffner et al. 2019).

The sources and potential uncertainties in the exposure assessment are described and systematized in 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 job-exposure matrix with object-, calendar-year- and activity-specific exposure), assignment error (assignment of values in job-exposure matrix 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. Generalization and assignment errors fall under the shared error type illustrated in Figure 3.1.

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 mine shafts) and parameter uncertainties, specifically uncertainties in the determination of parameters (*e.g.*, expert exposure 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 relatively minor (Deffner et al. 2019). Generalization error and parameter uncertainties 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 included in calculations. 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.

As described above, there are numerous sources of uncertainty associated with studies of underground miners that may affect risk estimates. NCRP (2012) stated that the BEIR VI report (NA/NRC 1999) found “that errors in exposure estimation, scarcity of data regarding miner exposure to other lung carcinogens (smoking, arsenic, silica, diesel exhaust) and temporal expressions of risk are likely to be the primary sources of uncertainty in lung cancer risk estimates arising from epidemiological studies of miners.”

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 absorbed dose to the thyroid tissue from inhalation of radioactive particles. Another example might be a dose reconstruction study conducted to estimate the dose to thyroid tissue of individuals who were exposed to radioactive iodines as children 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: In the Wismut analysis (Deffner et al. 2019), several time-specific uncertain parameters were evaluated including radon concentrations, equilibrium factors and working hours per year. Additionally, the Chernobyl thyroid screening studies (Likhtarov et al. 2014; Drozdovitch et al. 2015) included uncertainties involving the concentration and size distribution of radioactive particles (^{131}I) 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 versus 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 thyroid screening studies (Likhtarov et al. 2014; Drozdovitch et al. 2015), the concentration of the radioactive particles (^{131}I) in air at an evaluation location could range from 10 to 200 kBq m⁻³, depending on the distance from the release site. The size distribution of the particles could range from 0.1 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, considering both subjective and objective factors.

Example: Depending on the available measurements, a log-normal distribution might be used for the concentration of the radioactive particles (^{131}I) 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 evaluations, interdependencies can arise between two parameters when changes in one parameter influence changes in the other. This indicates that the values of the two parameters are not independent. Understanding these interdependencies is essential for accurate data analysis and interpretation. If sufficient information is available, these dependencies can be characterized and utilized to propagate uncertainty to estimates of organ absorbed dose.

Example: Interdependencies might be identified between the concentration of radioactive particles and weather conditions, as higher wind speeds can disperse the radioactive plume more widely, resulting in a lower concentration at a given location. Interdependencies might also be noted between the size distribution of particles and the assumed behavior of individuals in the exposed population. For instance, smaller particles can penetrate deeper into the lungs and affect the dose to specific lung regions. Additionally, documented adverse weather events may lead to significant interdependencies between wind-speed and population behavior.

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 representing the resulting degrees of belief for the assessment endpoint (an example is provided in Figure 4.1). In Section 4.4 numerical methods for error propagation are explored in greater depth.

Example: 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. Alternatively, a correlation matrix can be developed to characterize the shared error between cohort members.

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. Based on the results, recommendations can be made to proceed with the analysis or take 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.

Example: After a comprehensive analysis, the sources of uncertainty in the Wismut cohort were characterized and classified into major, medium and minor (Deffner et al. 2019).

4.4 Numerical Models for Error Propagation in Uncertainty Analysis

Various numerical methods can be used to perform an uncertainty analysis to overcome problems encountered with uncertainty propagation for complex models (NCRP 1996). While most methods require substantial computational resources, it is nevertheless important to evaluate and communicate this error for subsequent statistical analysis. Two examples have been included to illustrate methods of characterizing and communicating dosimetric uncertainty from reconstructions. The first example below employs a Monte Carlo analysis to determine the dosimetric probability density function for individuals where there is negligible shared error (IAEA 1989). The second example emphasizes the development of dose estimates for multiple realizations, which is a technique of particular value when dealing with the expectation of substantial shared errors (NCRP 2009b; Simon et al. 2015). These two examples account 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 the most applied numerical technique is Monte Carlo simulation, there are other approaches for error analysis including:

- deterministic 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 et al. 2003; de Wilde and Tian 2009); and
- nonprobabilistic methods, for example:
 - fuzzy sets and arithmetic (Puri et al. 1993; Cooke 2004; Baudrit et al. 2008); and
 - possibility theory (Dubois et al. 2001).