université PARIS-SACLAY

NNT : 2017SACLS532

UNIVERSITÉ PARIS SUD
Comprendre le monde, construire l'avenir

Approche hiérarchique bayésienne pour la prise en compte d'erreurs de mesure d'exposition chronique et à faibles doses aux rayonnements ionisants dans l'estimation du risque de cancers radio-induits. Application à une cohorte de mineurs d'uranium

Thèse de doctorat de l'Université Paris-Saclay
préparée à Paris Sud

École doctorale n°570 Santé publique
Spécialité de doctorat: Biostatistique

Thèse présentée et soutenue à Fontenay-aux-Roses, le 12 décembre 2017, par

Sabine Hoffmann

Composition du Jury :

Mme Margot Tirmarche Commissaire, Responsable de Recherche, Autorité de Sûreté Nucléaire	Présidente
M Jean-Michel Marin Professeur, Université de Montpellier	Rapporteur
Mme Nora Fenske Chargée de Recherche, Bundesamt für Strahlenschutz	Rapporteur
M Eric Parent Professeur, AgroParisTech	Examinateur
Mme Anne Thiébaut Chargée de Recherche, INSERM	Examinatrice
Mme Chantal Guihenneuc Professeur, Université Paris Descartes	Directrice de thèse
Mme Sophie Ancelet Chargée de Recherche, Institut de Radioprotection et de Sûreté Nucléaire	Encadrante de thèse

Thèse de doctorat

université PARIS-SACLAY | ÉCOLE DOCTORALE Santé Publique

Titre : Approche hiérarchique bayésienne pour la prise en compte d'erreurs de mesure d'exposition chronique et à faibles dose aux rayonnements ionisants dans l'estimation du risque de cancers radio-induits. Application à une cohorte de mineurs d'uranium.

Mots clés : Erreurs de mesure, modélisation hiérarchique, statistique bayésienne, cohorte professionnelle, radon, cancer du poumon

Résumé : En épidémiologie des rayonnements ionisants, les erreurs de mesure d'exposition et l'incertitude sur le calcul de la dose absorbée à l'organe constituent des sources d'incertitude importantes entrant en jeu dans la modélisation et l'estimation des relations dose-réponse d'intérêt. Celles-ci peuvent être de nature très complexes dans le cadre d'études de cohortes professionnelles et sont ainsi rarement prises en compte dans ce domaine. Pourtant, lorsque les erreurs de mesure d'exposition ne sont pas ou mal prises en compte, elles peuvent mener à des estimateurs de risque biaisés, à une perte de puissance statistique ainsi qu'à une déformation de ces relations dose-réponse. L'objectif de ce travail est de promouvoir l'utilisation de l'approche hiérarchique bayésienne pour la prise en compte explicite et simultanée des erreurs de mesure d'exposition et des incertitudes de dose intervenant dans les estimations de risques sanitaires radio-induits dans les études de cohortes professionnelles. Plus spécifiquement, des modèles hiérarchiques ont été proposés et inférés afin d'affiner l'estimation actuelle du risque de décès par cancer du poumon associé à une exposition chronique et à faibles doses au radon et ses descendants à vie courte à partir de données issues de la cohorte française des mineurs d'uranium. Ces modèles, connus pour leur souplesse et leur pertinence pour la prise en compte de sources d'incertitude multiples et complexes, sont basés sur une combinaison de sous-modèles probabilistes conditionnellement indépendants. Afin de quantifier l'impact de l'existence d'erreurs de mesure d'exposition partagées et non-partagées sur l'estimation du risque et sur la forme de la relation exposition-risque dans les études de cohortes professionnelles, une étude par simulations a été conduite dans laquelle différentes structures complexes mais réalistes d'erreurs de mesure (pouvant par ailleurs varier dans le temps) ont été considérées. Une élicitation de lois a priori reflétant l'incertitude relative au débit respiratoire - un paramètre important intervenant dans le calcul de la dose absorbée au poumon – a été conduite auprès de trois experts des conditions d'exposition dans les mines d'uranium française et des méthodes de combinaison de dires d'experts ont été mises en œuvre et comparées. Enfin, des algorithmes Monte-Carlo par Chaînes de Markov ont été implémentés sous Python pour mener l'inférence bayésiennes des différents modèles hiérarchiques proposés et ainsi, obtenir des estimations corrigées du risque de décès par cancer du poumon radio-induit dans la cohorte française des mineurs d'uranium.

Title : A Bayesian hierarchical approach to account for measurement error in the protracted low-dose exposure to ionizing radiation in the estimation of the risk of radiation-induced cancers. Application to a cohort of uranium miners.

Keywords : Measurement error, hierarchical modelling, Bayesian statistics, occupational cohort, radon, lung cancer

Abstract : In radiation epidemiology, exposure measurement error and uncertain input parameters in the calculation of absorbed organ doses are among the most important sources of uncertainty in the modelling of the health effects of ionising radiation. As the structures of exposure and dose uncertainty arising in occupational cohort studies may be complex, these uncertainty components are only rarely accounted for in this domain. However, when exposure measurement is not or only poorly accounted for, it may lead to biased risk estimates, a loss in statistical power and a distortion of the exposure-response relationship. The aim of this work was to promote the use of the Bayesian hierarchical approach to account for exposure and dose uncertainty in the estimation of the health effects associated with exposure to ionising radiation in occupational cohorts. More precisely, we proposed several hierarchical models and conducted Bayesian inference for these models in order to obtain corrected risk estimates on the association between exposure to radon and its decay products and lung cancer mortality in the French cohort of uranium miners. The hierarchical appraoch, which is based on the combination of sub-models that are linked via conditional independence assumptions, provides a flexible and coherent framework for the modelling of complex phenomena which may be prone to multiple sources of uncertainty. In order to compare the effects of shared and unshared exposure uncertainty on risk estimation and on the exposure-response relationship we conducted a simulation study in which we supposed complex and potentially time-varying error structures that are likely to arise in an occupational cohort study. We elicited informative prior distributions for average breathing rate, which is an important input parameter in the calculation of absorbed lung dose, based on the knowledge of three experts on the conditions in French uranium mines. In this context, we implemented and compared three approaches for the combination of expert opinion. Finally, Bayesian inference for the different hierarchical models was conducted via a Markov chain Monte Carlo algorithm implemented in Python to obtain corrected risk estimates on the lung cancer mortality in the French cohort of uranium miners associated with exposure to radon and its progeny.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Acknowledgements

Puisqu'on ne peut pas être universel et savoir tout ce qu'on peut savoir sur tout, il faut savoir un peu de tout. Car il est bien plus beau de savoir quelque chose de tout que de savoir tout d'une chose; cette universalité est la plus belle. -
Pensées (1670) de Blaise Pascal.

I would like to thank my two supervisors Chantal Guihenneuc and Sophie Ancelet: First of all, for giving me the opportunity to work on the subject of this PhD thesis that I enjoyed working on very much. I have now been waiting for three years to get tired of measurement error, Bayesian statistics and uranium miners, but even as I write these lines and as I send you this manuscript I continue waiting. I know that this is in great part due to all the freedom and trust you gave me during the last three years. However, above all it is due to your humour, your enthusiasm, your patience and your passion for your work. I consider myself to be very fortunate to have been able to work with you for the last years and I could not imagine a better team to supervise a PhD.
I would also like to thank my rapporteurs, Nora Fenske and Jean-Michel Marin, for having accepted to take time from their busy schedules to review this work. I would like to thank Eric Parent, Anne Thiébaut and Margot Tirmarche for having accepted to be examinateurs.
I owe a lot to the whole laboratory of epidemiology of the IRSN and in particular Dominique Laurier and Klervi Leraud for being always open to questions about the French cohort of uranium miners as well as for their great leadership of this laboratory. Thank your for your patience and for your openness. A particular thanks to Estelle Rage, Sylvaine Caer-Lorho and David Richardson for all their answers to my questions about uranium miners. I would like to thank Marion Belloni for spending three afternoons with me in front of my computer to search an exceptionally stupid error in my Python code. I would also like to thank Lucie Fournier. I benefited a lot from your advice, in particular concerning all the administrative tasks that come along with a Phd in France and I feel like you always managed to finish before I even started thinking about them. I would also like to thank Guillaume Wyart and Segolene Bouet for making this last year of work as joyful, entertaining and productive as it was and in particular for all your advice on how to prepare "des topinambours", des "courges shaghetti", du "chou câle" and all the other vegetables that arrived in my "panier bio" every week.
As I wrote large parts of this PhD thesis on long train and bus rides, I would like to thank SNCF, Deutsche Bahn, Thalys, FlixBus, Renfe and Comboios de Portugal for their numerous delays that gave me even more time for my writing. On a more serious note, I am grateful for the European Union that both enabled me to do an Erasmus semester abroad and facilitated my decision to stay in France to continue my studies in Rennes and in Paris. I am very grateful for

my friends, no matter if they are living in France, Germany, Belgium or Colombia. Thank you for your advice and your patience with me in the last three years, but also for distracting me from my work. I would like to thank my parents for their generosity in allowing me to enjoy all the education I wished for, making it possible for me to obtain three university degrees in two different countries, which all helped me to some degree in the work on this PhD thesis. Thank you for always believing in me in spite of all evidence to the contrary. I do not know whether I would do the same for a child but I'll try to remember your generosity if I should ever be faced with such a decision. I would also like to thank my brother Michael for his infinite patience in helping me rediscover mathematics when working on my weekly homework in linear algebra and analysis in my first year of statistics. Without you there are good chances that I would have become a psychotherapist after all, spending my days with depressed patients rather than with cheery data, and I am convinced that we are all better off with this choice. Thank you Ute for your moral and material support in this PhD thesis but also for always reminding me that there are more important things in the world than my work. I thank my uncle Seppi for his patience and for all his passion for the beta distribution and for mathematics in general.
Thank you Ronan, simply for being who you are every day.

Valorisation scientifique de la thèse

Publications avec comité de lecture

Articles issus de la thèse

Hoffmann S, Rage E, Laurier D, Laroche P, Guihenneuc C, Ancelet S. Accounting for Berkson and classical measurement error in radon exposure using a Bayesian structural approach in the analysis of lung cancer mortality in the French cohort of uranium miners. Radiation Research. 2017; 187(2), 196-209.

Hoffmann S, Rage E, Laurier D, Guihenneuc C, Ancelet S. Shared and unshared measurement error in occupational cohort studies and their effects on statistical inference in proportional hazards models. PLOS ONE. *Accepted with minor revisions.*

Hoffmann S, Guihenneuc C, Ancelet S. A cautionary comment on the generation of Berkson error in epidemiological studies. Radiation & Environmental Biophysics *Accepted with minor revisions.*

Autres articles

Hoffmann S, Sobotzki C. Response to Avoidance of sun exposure as a risk factor for major causes of death: a competing risk analysis of the Melanoma in Southern Sweden cohort. Journal of Internal Medicine. 2017; 281(6), 622-623.

Durand T, Bompaire F, Hoffmann S, Léger I, Jego V, Baruteau M, Delgadillo D, Taillia H, Richard D. Episodic memory impairments in primary brain tumour patients. Archives of Clinical Neuropsychology. *Accepted.*

Communications orales

Hoffmann S, Ancelet S, Laroche P, Guihenneuc C. **Using a structural Bayesian approach to account for measurement error: An application to radiation epidemiology.** *47ème Journées de statistique de la SFdS, Lille (France), June 2015.*

Hoffmann S, Guihenneuc C, Ancelet S. **Approche hirarchique baysienne pour la prise en**

compte d'erreurs de mesure d'expositions chroniques aux rayonnements ionisants dans l'estimation du risque de cancers radio-induits. Application une cohorte de mineurs d'uranium. *AppliBUGS. Paris, June 2015.*

Hoffmann S, Guihenneuc C, Laroche P, Ancelet S. **A Bayesian structural approach to account for uncertainty in exposure assessment in a French cohort of uranium miners.** *36th Annual Conference of the International Society for Clinical Biostatistics, Utrecht (Netherlands), August 2015*

Hoffmann S, Ancelet S, Laroche P, Guihenneuc C. **Accounting for exposure uncertainty and effect modification using a Bayesian structural approach: Radon exposure and lung cancer mortality in a prospective cohort of uranium miners.** *XXVIIth International Biometric Conference, Victoria (Canada), July 2016.*

Hoffmann S, Guihenneuc C, Laroche P, Ancelet S. **Modeling effect modification and exposure uncertainty in the association between lung cancer mortality and radon exposure in a cohort of uranium miners via a Bayesian hierarchical approach.** *ISEE, Rome (Italy), September 2016.*

Hoffmann S, Ancelet S, Laroche P, Guihenneuc C. **Accounting for Berkson and classical measurement error in radiation epidemiology studies using a Bayesian structural approach. Application to the French cohort of uranium miners.** *EPICOH, Barcelona (Spain), September 2016.*

Ancelet S, Hoffmann S. **Internal doses and uncertainties: the point of view of the epidemiologist.** *Radiation Protection Week, Oxford (United Kingdom), September 2016.*

Hoffmann S, Guihenneuc C, Ancelet S. **The effects of shared and unshared exposure uncertainty on risk estimation via proportional hazards models in occupational cohorts.** *49me Journes de Statistique de la SFdS, Avignon, June 2017*

Hoffmann S, Guihenneuc C, Ancelet S. **Deriving informative prior distributions by the combination of the opinions of several experts in a Bayesian hierarchical approach.** *An application in radiation epidemiology. AppliBUGS. Paris, May 2017.*

Hoffmann S, Ancelet S, Guihenneuc C. **Elicitation of the knowledge of several experts to derive informative prior distributions on biological parameters in radiation epidemiology.** *38th Annual Conference of the International Society for Clinical Biostatistics, Vigo (Spain), July 2017*

Communications affichées

Hoffmann S, Guihenneuc C, Ancelet S. **Shared and unshared exposure uncertainty in proportional hazards models: Effects on risk estimation and shape of the exposure-response curve** *38th Annual Conference of the International Society for Clinical Biostatistics, Vigo (Spain), July 2017*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Exposure measurement error poses one of the most important threats to the validity of statistical inference in epidemiological studies [3, 4, 5]. It arises whenever an exposure cannot be measured accurately, but only an imperfect surrogate exposure measurement is available. When it is not or only poorly accounted for, exposure measurement error can lead to biased risk estimates, a loss in statistical power and to a distortion of the exposure-risk relationship [6, 7, 8]. Despite these deleterious consequences and despite its ubiquity in observational research [9, 10, 11], exposure measurement error is only rarely accounted for in the estimation of risk coefficients in epidemiological studies [12, 13]. One of the main reasons why measurement error is often discussed, but rarely accounted for in epidemiology, may be that classical methods which are routinely used to correct for measurement error, like simulation extrapolation or regression calibration, lack the flexibility to account for complex patterns of exposure uncertainty. Methods to account for more complex error structures, on the other hand, are neither easy to implement nor readily available in standard software packages.

In occupational cohort studies, for instance, one is commonly interested in the association between the time until the diagnosis or the time until the death by a certain disease and cumulative exposure to a specific chemical or physical agent. In this situation, cumulative exposure is ongoing and time-dependent, rather than being a fixed point exposure [14], which could be determined at study entry. Owing to this time-dependent nature, the exposure history of workers may be collected using different strategies according to the period of exposure. Thereby, the changes in the methods of exposure assessment can create rather complex patterns of exposure uncertainty, where the type and magnitude of measurement error can vary over time. For recent exposure periods, it is common to dispose of a method of prospective, and possibly individual, exposure monitoring in these studies. In these periods, technical advances in measurement devices may imply more and more precise measures of exposure, which can translate into a decrease in the variance of measurement error over time. On the other hand, it is very rare to dispose of a systematic exposure assessment for the earliest years of exposure in an occupational cohort study, simply because there was often a lack of awareness of the deleterious health consequences of the exposure of interest in these years. As a consequence, the exposure values received in this period often have to be reconstructed retrospectively. As it is virtually impossible to reconstruct the exposure values for each individual worker in a retrospective fashion, one usually has to estimate the exposure levels for different job categories for the earliest exposure periods. In this context,

the same exposure level is affected for all workers in a given job category. If there is an error in the estimation of this common exposure level, it will affect all workers in that job category in the same way. In other words, a retrospective exposure reconstruction may give rise to error components that are shared among several workers. Additionally, when cumulative exposure is modelled as a time-dependent variable and a method of group-level exposure estimation is used, individual job conditions and worker practices may create statistical dependencies between measurement errors in the exposure history of a worker [15, 16]. These dependencies can be described by a measurement error component that is shared within workers, that is, an exposure uncertainty component that affects all exposure values received by a worker in the same way. While there is a vast literature on the topic of unshared measurement error, statistical methods for the treatment of shared error components are scarce and the effects of these error components on statistical inference remain largely unknown. At the same time, it is likely that the complex patterns of measurement error that arise in occupational cohort studies pose serious threats to the validity of statistical inference in these studies. In particular, they may induce systematic distortions of the exposure-response relationship, because both average exposure and the magnitude of measurement error in an occupational cohort tend to decrease over time [17, 14].

In occupational cohort studies in radiation epidemiology, the problem of measurement error may be further exacerbated by the fact that the health effects of radiation are associated with radiation dose, rather than with radiation exposure [18, 19]. The values of radiation dose do not only depend on the exposure to radioactive material, but also on the exposure conditions. In particular, when we are interested in doses to internal organs that can be due to the inhalation or the ingestion of radioactive material, these doses do not only depend on the exposure to this radioactive material but also on the rate at which this material was inhaled or ingested. The different rates at which radioactive material is inhaled, for instance, can be modelled by parameters like average breathing rate, which are typically introduced as additional input parameters intervening in dose calculation. As the input parameters in dose calculation which modify the relation between radon exposure and radon dose are rarely measured in occupational cohorts, the calculation of radiation doses is uncertain when estimating the health effects of radiation exposure in occupational cohort studies. A current methodological challenge in radiation epidemiology is therefore to be able to account for dose uncertainty in risk estimation.

Finally, the field of radiation epidemiology is particularly interested in the estimation of the effects of chronic low-dose radiation on health. If the effects of exposures received at high to moderate doses, which occurred for instance in the earliest exposure periods in most occupational cohort studies or in the Life Span Study[1], have to be extrapolated to low-dose and low-dose rate exposures, the adequacy of this extrapolation depends on an accurate estimation of the dose-response relationship. If, on the other hand, the effects of chronic low-dose radiation are directly estimated on populations that are exposed at low-dose and low-dose rates, there may be only a weak association between radiation exposure and its health effects. In this situation, the potential consequences of measurement error, in particular bias in risk estimates and loss

[1]The Life Span Study, which consists of the survivors of the atomic bombings of Hiroshima and Nagasaki, is an important source of information on the health effects of radiation exposure. While it played, and continues to play, an important role in the establishment of radiation protection guidelines, it primarily provides information on the health effects observed on a population that received an instantaneous exposure to ionizing radiation at high to moderate doses.

of statistical power, may lead to erroneous conclusions, as these studies may fail to detect the already weak associations between a chronic low-dose exposure to ionising radiation and the health outcomes of interest, or maybe even worse, mistakenly detect significant associations where there are none.

In summary, it is important to account for exposure and dose uncertainty in studies in radiation epidemiology in order to obtain reliable estimates concerning the health effects of radiation exposure and to avoid distortions of the dose-response relationship. However, this task is very challenging and requires advanced modelling and statistical methods. Indeed, we may be faced with complex error structures including changes in the type and magnitude of measurement error over time, components of measurement error that may be shared between individuals, within individuals or both and uncertainty in other parameters intervening in dosimetric modelling.

This work will be concerned with the analysis of the association between exposure to radon and its progeny (denoted radon hereafter) and lung cancer mortality in the French cohort of uranium miners. Radon is a noble and radioactive gas that naturally occurs in the decay of uranium-238, which is omnipresent in soils and rocks. The major source of exposure to radiation in most countries is due to natural background radiation and the most important component of this background radiation is exposure to radon [20]. Radon was classified as a pulmonary carcinogen in humans by the International Agency for Research on Cancer in 1988 [21]. It is considered to be the second leading cause of lung cancer after smoking [22, 20], causing around 2% of cancer deaths in Europe every year [23]. Radon can be seen as the biggest geological hazard to humans worldwide, causing more deaths than earthquakes, hurricanes and tsunamis together. A variety of techniques to reduce radon concentration in existing and new buildings [22] is available, including underfloor ventilation and the installation of radon proof barriers at ground level [23]. However, the exact knowledge of the risk associated with radon exposure is necessary to determine the cost effectiveness of possible radon reduction policies [22].

Much of the evidence on the health effects associated with radon originates from epidemiological studies on cohorts of occupationally exposed underground miners. The problem of exposure measurement error in the association between radon exposure and lung cancer mortality has attracted considerable attention [24, 25, 26, 27, 28, 29, 30, 31, 23, 32, 33, 34, 35, 36]. However, previous studies addressing the problem of measurement error in radon exposure in cohorts of uranium miners have made a number of simplifying assumptions. Bender et al. (2005) [31] and Küchenhoff et al. (2007) [32], who studied the impact of measurement error in radon exposure on risk estimation in the German cohort of uranium miners, neglected the time-varying nature of cumulative exposure by assuming that the sum of the annual exposure values received during the entire working career of a miner is known at study entry. While it is common to make this assumption when treating the problem of exposure measurement error in occupational cohort studies [17, 14], it impedes the modelling of measurement error on its natural level of occurrence, namely on the weekly, monthly or annual exposure values, rather than on the sum of these values. Moreover, they assumed that errors arising in an exposure assessment via job-exposure matrices can be described by unshared measurement error, thereby neglecting the potential of individual job characteristics or worker practices to lead to an error component shared within workers. Likewise, Allodji et al. (2012) [34, 35], who studied the impact of exposure measurement error when estimating the association between radon exposure and lung cancer mortality

in the French cohort of uranium miners, assumed unshared measurement error. Moreover, he promoted the use of simulation extrapolation and regression calibration, which arguably lack the flexibility to account for complex structures of measurement error that typically arise in occupational cohort studies.
The main aim of this work was to promote the use of the Bayesian hierarchical approach to account for exposure and dose uncertainty in radiation epidemiology. More specifically, our aim was to build different hierarchical models accounting for exposure and dose uncertainty in radon exposure in the French cohort of uranium miners and to conduct Bayesian inference for these models in order to obtain corrected estimates on the lung cancer risk that is associated with radon exposure in this cohort. The hierarchical approach, which is based on the combination of sub-models via conditional independence assumptions, provides a coherent framework for the treatment of complex phenomena which may be prone to multiple sources of uncertainty. The choice to conduct Bayesian inference to fit the proposed hierarchical models is not only based on its coherence when it comes to the estimation of complex probability models that are based on conditional independent sub-models, but also on its flexibility and the possibility to integrate expert knowledge through the elicitation of informative prior distributions. While a number of sensitivity analyses have studied the impact of dosimetric uncertainties in the association between radon exposure and lung cancer mortality [2, 37, 38, 39], to our knowledge this is the first time that shared exposure uncertainties and dosimetric uncertainties are accounted for when studying the association between radon exposure and lung cancer mortality in a cohort of uranium miners.

The outline of this manuscript is as follows:

Chapter 2 gives a short introduction to ionising radiation, followed by a more detailed presentation of radiation exposure in cohorts of uranium miners, dose estimation for radon exposure and results on the association between radon exposure and lung cancer mortality.

Chapter 3 provides an overview of the literature on different types of measurement error, their potential effects on risk estimates in epidemiological studies and of standard approaches to correct for these effects.

Chapter 4 gives a brief overview of both the Bayesian framework to statistical inference and the Bayesian hierarchical approach with a particular focus on its ability to account for exposure uncertainty in epidemiological studies and on the possibility to elicit prior distributions on unknown parameters based on expert knowledge.

Chapter 5 describes the different methods we used in this work and the simulation studies we conducted. It presents the different hierarchical models that were proposed to account for exposure and dose uncertainty in the French cohort of uranium miners and the algorithms that were implemented to conduct Bayesian inference for these models. It presents the methods that we proposed to derive informative prior distributions for the most important input parameters in the dose model. In particular, it will present an elicitation approach that we developed to assign

a prior distribution on a specific input parameter. Finally, it describes two simulation studies that were conducted in order to study the effects of different types of measurement error on statistical inference in occupational cohort studies when measurement error is not accounted for.

Chapter 6 presents the results we obtained in the two simulations studies that we conducted in the context of this work, as well as the informative prior distributions that were obtained when eliciting the opinion of three experts on the exposure conditions in French uranium mines. Moreover, it provides the risk estimates that were obtained for lung cancer mortality when accounting for dose uncertainty and different types of shared and unshared exposure uncertainty in the French cohort of uranium miners.

Chapter 7 presents a general discussion of the strengths and limitations of this work. Moreover, it discusses possible implications for radiation protection and the potential of the Bayesian hierarchical approach to account for exposure and dose uncertainty in radiation epidemiology.

Chapter 2

Radon exposure and dose estimation in cohorts of uranium miners

The high mortality rate associated with working in underground mines was documented as early as the 15th century in the Ore mountains in Central Europe [40, 41], long before the discovery of radioactivity. Around 1530, the famous Swiss physician and scientist Philippus Aureolus Theophrastus Bombastus von Hohenheim (also known as Paracelsus) studied the health conditions of underground miners in the Ore mountains and reported that many of them died of lung disease [41, 42]. In 1879, Harting and Hesse identified the cause of this lung disease as lung cancer and it became known as the 'Schneeberger Bergkrankheit" [43, 44] (German for "mountain disease from the snow mountain"). After this discovery, it took 70 years to recognise that this high lung cancer mortality might be caused by the inhalation of short-lived radon progeny [45, 40] and almost another 40 years before the International Agency of Research on Cancer classified radon as known pulmonary carcinogen in humans [21]. While the lung cancer mortality rates that occurred in the mines in Schneeberg and Joachimmsthal in the Ore mountains during the 19th century were estimated to be around 50% [46] and up to 75% [41], the conditions in mines nowadays generally result in an annual exposure to radon progeny comparable to the exposure level experienced in a large number of houses [47]. This chapter gives a short introduction to ionising radiation, followed by a more detailed presentation of radiation exposure in cohorts of uranium miners, dose estimation for radon exposure and results on the association between radon exposure and lung cancer mortality.

2.1 Ionising radiation

While the Ore mountains are relatively unknown, both the term Dollar (derived from Thaler as a reference to Joachimsthal), which is used for more than twenty currencies nowadays, and the discovery of radioactivity can be traced back to this region, which separates Saxony and Bohemia [42, 48].
Martin Heinrich Klaproth discovered uranium in pitchblende[1] from this area in 1789 [42]. Over

[1]Pitchblende, from German "Pechblende", is a uranium-rich mineral that may contain copper, bismuth, barium, lead and rare earth elements [42]. "Pech" means bad luck in German, thereby referring to the fact that it

a century later, in 1896, Henri Becquerel conducted experiments on uranium salts from the Ore mountains and famously discovered evidence of radioactivity by a happy accident. He was looking for X-ray-like radiation, which had been discovered by Wilhelm Conrad Röntgen one year earlier [49]. As an expert in fluorescence, he supposed that the radiation emitted by uranium salts needed to be activated by sunlight [48, 42]. He exposed the salts to the sun and placed them on photographic plates to detect radiation. As the sky of Paris was cloudy on the 26th and 27th of February 1896, he postponed his experiment and kept the salt and the photographic plates in a drawer, only to discover to his astonishment that the radiation seemed to arise spontaneously from the uranium salts [48].
In his serendipity, and without understanding its origin, he had discovered ionising radiation, that is radiation that can interact with matter and free electrons by transferring the energy it carries to an atom or a molecule [50]. Fascinated by this discovery, Marie Sklodowaska-Curie decided to select this topic for her PhD thesis [51, 42]. In 1898, she and her husband examined pitchblende from the Joachimsthal mines and reached the conclusion that unknown compounds, even more radioactive than uranium, were present in the ores [51].

2.1.1 Radioactive decay

The origins of ionising radiation can be found in the atomic nucleus, which was discovered by Ernest Rutherford in 1911[2] [52]. According to the currently accepted view of the atomic structure [53], the atomic nucleus is built of subatomic particles, which are called nucleons. These nucleons are surrounded by an electron cloud, as illustrated in Figure 2.1, where the position of each electron can be described in terms of a probability distribution. There are two types

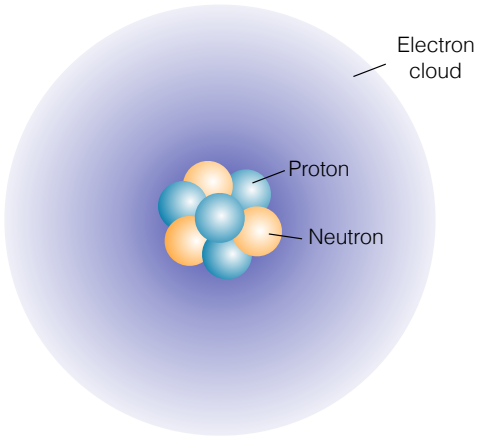

Figure 2.1: The modern atomic model

of nucleons: protons and neutrons [54]. Protons have a positive electric charge and neutrons have no charge [50], resulting in an overall positive charge of the nucleus, while the electrons

was a sign of bad luck for the silver miners in Saxony to find this uranium rich ore, as it meant that the silver deposits were running out [42].

[2]Rutherford is considered to be the founder of nuclear physics [52]. Ironically, Rutherford was awarded the Nobel prize in Chemistry in 1908 but never received the Nobel Prize in Physics. Marie Curie, on the other hand, received the Nobel Prize in Physics in 1903 and the Nobel Prize in Chemistry in 1911. She is the only person having received a Nobel Prize in two different sciences as well as the first woman to receive a Nobel Prize.

are negatively charged [49]. Based on the electromagnetic interactions between the positively charged protons in the nucleus, one would expect these protons to repel each other, resulting in extremely unstable nuclei. However, this electromagnetic force is balanced by what is called the strong and the weak nuclear force. These forces are of shorter range than the electromagnetic force, but they are in general sufficiently strong to overcome the electromagnetic repulsion between protons [52].
The nucleus is commonly characterised by its number of neutrons and its number of protons. A nucleus containing Z protons and N neutrons is expressed as ${}^{A}_{Z}X$, where A is the sum of Z and N, i.e., the total number of nucleons or the atomic mass number and X is the chemical symbol of the element with Z protons [55]. Each element is uniquely defined by its number of protons. Atoms that have the same number of protons, but a different number of neutrons are called isotopes. Isotopes of the same element have usually very similar chemical properties but they may exhibit very different nuclear properties.
These nuclear properties mainly depend on the proton-neutron ratio of a nucleus and its mass number [50]. In particular, if a nucleus is too heavy or if the ratio between protons and neutrons in the nucleus is not optimal it can be unstable. In the following, we will refer to the unstable isotopes of an element as radioisotopes. Radioisotopes can spontaneously change its configuration by emitting α- or β- particles or γ-radiation. This process is known as radioactive decay.

Alpha decay

Alpha decay, denoted α-decay in the following, mainly occurs in heavy radioisotopes [55]. In this form of radioactive decay, the nucleus emits an α-particle, which is equivalent to a helium nucleus, consisting of two protons and two neutrons, carrying a double positive charge (see Figure 2.2). By emitting an α-particle, the nucleus changes its atomic mass number by four and

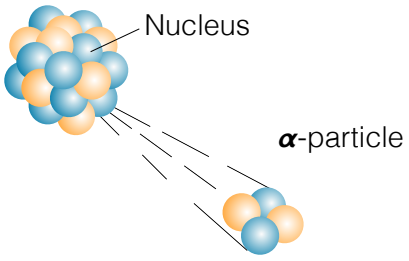

Figure 2.2: Alpha decay

its atomic number by two. As α-particles are relatively heavy and charged, they tend to interact with other atoms very quickly and therefore have a very limited ability to penetrate matter. For instance, they can be absorbed by a piece of paper or the outer layer of the skin and travel only a few centimetres in air. Consequently, this type of radiation only presents a radiation hazard if radioisotopes enter the body through the inhalation or the ingestion of radioactive material.

Beta decay

Beta decay, denoted β-decay in the following, is a phenomenon that changes the ratio between protons and neutrons in a nucleus. In this type of decay, either an electron (β^- decay) or its antiparticle, a positron (β^+ decay) [55], is emitted from the nucleus, transforming a neutron into a proton or vice versa [50]. As electrons and positrons are less heavy and carry less charge than α-particles, they react less readily with matter, resulting in a maximum path length of beta radiation that is larger than for α-radiation with some milimeters in tissue and a few meters in air [50].

Gamma radiation

After the emission of an α- or a β-particle, a nucleus may remain in an excited state, characterised by a high energy level. In this situation, a common phenomenon is the emission of a Gamma-ray to further decay to the ground state [55]. Gamma radiation, denoted γ-radiation in the following, is highly energetic electromagnetic radiation. γ-rays have zero rest mass and zero electric charge, they are more penetrating than α- and β-radiation [50].

2.2 Radon exposure in cohorts of uranium miners

In the context of their work, underground miners are exposed to α-, β- as well as γ- radiation, as all three types of radiation arise in the decay of uranium. Uranium is a chemical element, which is ubiquitous in the earth's crust with low levels of uranium that can be found in most rocks, soils and water [56]. With 92 protons and between 141 and 146 neutrons, the nucleus of uranium is too heavy to be stable and therefore all isotopes of uranium are radioactive. The most common isotope of uranium found in nature is uranium-238 ($^{238}_{92}$U). This isotope makes up over 99% of natural uranium and it has a half-life[3] of 4.468 billion years. Due to the ubiquity of uranium, all underground miners, regardless of the type of mine, are exposed to the daughter products of uranium, but the uranium concentrations in the ore mined in uranium mines is typically higher than in other types of mines. We will see in the following that radon plays a central role when it comes to the radiation exposure of uranium miners.

2.2.1 The uranium decay series

The uranium-238 decay series is one of the three decay series that occur naturally on Earth [57]. Figure 2.3 shows the radioisotopes in the uranium decay series from uranium-238 to radon-222, their most probable type of radioactive decay and their half-life. Uranium, thorium, protactinium and radium decay by emitting α- or β-radiation. As α- and β-decay are often followed by γ-radiation, the decay from uranium-238 to radon-222 may be accompanied by γ-radiation. Since uranium, thorium, protactinium and radium are all in a solid state under the conditions which are typically encountered in underground mines and given the very short range of α- and β-radiation, the α- and β-radiation emitted by these radioisotopes does not

[3] The half-life of a radioisotope is the average time it takes half of the nuclei of this radioisotope to decay [50]

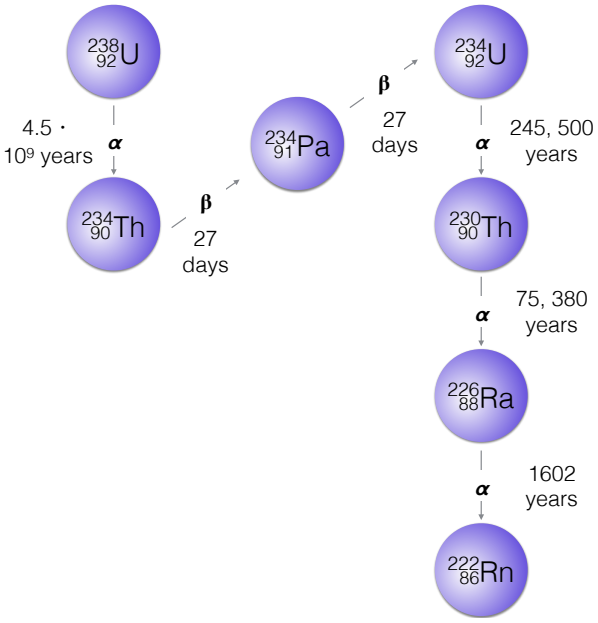

Figure 2.3: Radioactive decay chain from uranium-238 to radon-222. Uranium-238 decays into thorium-234 ($^{234}_{90}$Th), protactinium-234 ($^{234}_{91}$Pa), uranium-234 ($^{234}_{92}$U), thorium-230 ($^{230}_{90}$Th), radium-226 ($^{226}_{88}$Ra) and radon-222 ($^{222}_{86}$Rn).

pose a noteworthy exposure to ionising radiation[4], whereas it can cause external exposure to γ-radiation. Contrary to the other daughter products of uranium-238, radon-222 ($^{222}_{86}$Rn) is a

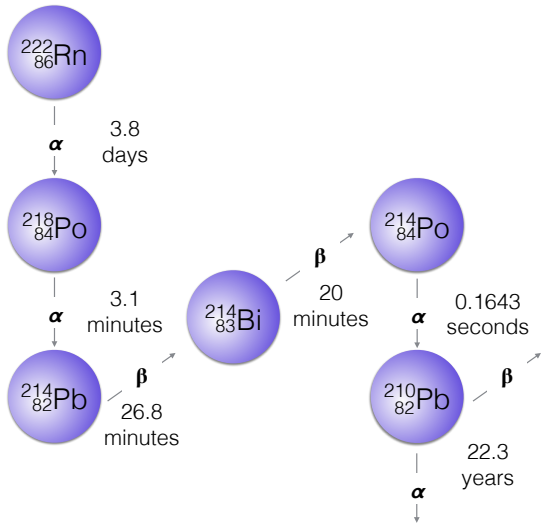

Figure 2.4: Radioactive decay chain from radon-222 to lead-210. Radon decays into polonium-218 ($^{218}_{84}$Po), lead-214 ($^{214}_{82}$Pb), bismuth-214 ($^{214}_{83}$), polonium-210 ($^{214}_{84}$Po), lead-210 ($^{210}_{82}$Pb) and long-lived radionuclides, which are not shown.

noble gas. As a gas, it can be inhaled by the workers in a mine. However, as a noble gas, radon has a very low chemical reactivity [58], making it colourless, tasteless and odourless. Due to

[4]Except when these radioisotopes are inhaled, for instance in the dust present in uranium mines.

this very low chemical reactivity, when inhaled, most of the gas will be subsequently exhaled [59]. At the same time, radon-222 decays into a series of short-lived radioisotopes, which are again solid, shown in Figure 2.4. These short-lived radioisotopes can form clusters, attach to ambient aerosol particles in the atmosphere [60] and, when inhaled, deposit in the lung where there can decay by emitting α-, β- and γ- radiation. In particular, polonium-218 (^{218}Po) and lead-214 (^{214}Po) can give rise to high lung doses from α-particle radiation [61]. To be precise, it is therefore not so much radon exposure, which is hazardous to humans and lung carcinogenic, but rather the exposure to radon decay products, also referred to as radon progeny. For sake of simplicity and because it is common in the literature to do so, we will discuss radon exposure in cohorts of uranium miners in the following, but it is important to note that we generally refer to radon progeny.

2.2.2 International cohorts of uranium miners

In order to study the association between radon exposure and lung cancer mortality, a number of cohorts of underground miners have been initiated and followed, including nine cohorts of uranium miners: the Czech cohort, the French cohort, the Radium Hill cohort (Australia), three Canadian cohorts (the Ontario, the Beaverlodge and the Port Radium Cohort), two US cohorts (the Colorado Plateau and the New Mexico cohort) and a German cohort (called the Wismut cohort). Lubin et al. (1995) [62] studied the association between radon exposure and lung cancer mortality in the pooled data of eight of these nine cohorts (excluding the Wismut cohort, which was established after the other eight cohorts). Moreover, in their analyses, the authors included data on three occupational cohorts of radon exposed miners who were not uranium miners: tin miners in China, iron miners in Sweden and fluorspar miners in Canada. The results of this pooled data are presented in detail in the BEIR VI report [63].

The French cohort of uranium miners

The French cohort of uranium miners is a prospective cohort that was initiated in the early 1980s. The current cohort consists of 5086 males, who were employed as uranium miners for at least one year at the group Commissariat à l'Énergie Atomique - COmpagnie GÉnérale des MAtières nucléaires (CEA-COGEMA) between 1946 and 1990. The follow-up began in 1946 and ended on the 31^{st} December 2007. The date of cohort entry was defined as the date of first employment as uranium miner plus one year. The date of cohort exit was defined as the earliest among the following: date of death, date of loss to follow-up, date of 85^{th} birthday and 31^{st} December 2007. Information on vital status was obtained via the Répertoire National d'Identification des Personnes Physiques (RNIPP). Causes of death were recorded both through the follow-up by the occupational medicine department and by the information available through the French National Mortality Database, which contains all the information on death certificates in France. For deaths occurring between 1946 and 1968, the information concerning the cause of death came from the occupational medicine department, while after 1990, all information came from the National Mortality Database. For deaths between 1968 and 1990, information from both sources was available [64]. Miners were censored at the age of 85 years because the determination of the exact death cause can become imprecise after this age [65]. The cause of

death was coded according to the International Classification of Disease (ICD) code in use at the time of death. Table 2.1 and 2.2 summarise the vital status and the main characteristics of

Table 2.1: Vital status in the French cohort of uranium miners on the 31^{st} December 2007.

Vital status	N (%)
alive $<$ 85 years	2924 (57.5%)
alive $\geq$ 85 years	187 (3.7%)
dead (lung cancer)	211 (4.2%)
dead (other cause)	1724 (33.9%)
lost to follow-up	40 (0.8%)

Table 2.2: Main characteristics of the French cohort of uranium miners on the 31^{st} of December 2007.

	Mean	Standard deviation	(Min, Max)
Age at study entry (in years)	28.8	7.7	(16.0, 68.4)
Duration of follow-up (in years)	35.4	12.5	(0.1, 61.0)
Duration of employment (in years)	17.0	10.6	(1.0, 43.3)
Cumulated radon exposure[a] (in WLM)	36.1	71.2	(0.003, 960.1)
Duration of radon exposure (in years)	11.8	8.4	(1.0, 37.0)
Age at first radon exposure (in years)	29.6	7.8	(15.2, 64.0)

[a]Only calculated on exposed miners.

the French cohort of uranium miners on the 31^{st} of the December 2007. Overall, 2924 (57.5%) miners were still alive at the end of the follow-up, 1935 (38.0%) had died, 40 (0.8%) were lost to follow-up and 187 (3.7%) had reached the age of 85. Cause-of-death information was available for 1876 deaths (97.0%). Miners started working at a mean age of 28.8 years and worked for an average duration of 17.0 years. The follow-up was relatively long with a mean duration of 35.4 years. Of 5086 miners, 4133 were exposed to radon, and the average cumulated exposure among exposed miners was 36.1 working level months[5] (WLM), respectively. The mean age at first exposure to radon was 29.6 years, and the average duration of radon exposure was 11.8 years.

2.2.3 Methods of radon exposure estimation in cohorts of uranium miners

Measuring radon exposure is an error-prone process, not only because of technical or human error, but also because radon levels are inhomogeneous in both space and time, while measurement methods tend to be in fixed locations at fixed times. Moreover, when modern uranium mining began in the late 1940s [67], there was little thought for the protection of the environment and

[5]The Working Level Month (WLM) is a historical unit to express exposure to radon progeny in cohorts of uranium miners. One working level (WL) is equivalent to any combination of short-lived radon progeny in one litre of air that results in the ultimate emission of $13 \cdot 10^{11}$ electron volt of energy from alpha particles. The unit of working level month is defined as the product of the accumulated exposure to radon decay products (expressed in WL) and one working month (defined as 170 hours) [66].

the protection of workers and therefore there was no systematic exposure assessment in uranium mines for either radon or its decay products [33, 68]. Most of the exposure values received in these early years of uranium mining therefore had to be reconstructed for the purpose of the epidemiological studies conducted on the international cohorts of uranium miners.
In most cohorts, exposure estimates were mainly based on three different techniques, which will be described in more detail in this section. In the earliest years, exposure values were frequently obtained by retrospective exposure reconstruction, followed by area sampling methods in more recent years. Moreover, individual exposure measurements were obtained via personal dosimetry in the most recent years of exposure in the French cohort of uranium miners [33] and in a Czech mine, which is still in operation today [69]. While exposure estimates became more and more precise over time, there was a gradual decrease in the radon levels received by uranium miners in most countries due to technical improvements and to the introduction of radiation protection standards [46]. The different methods of exposure assessment and the gradual decrease in radon levels can be seen in Figure 2.5, which shows the annual radon exposure of exposed miners in the French cohort of uranium miners.

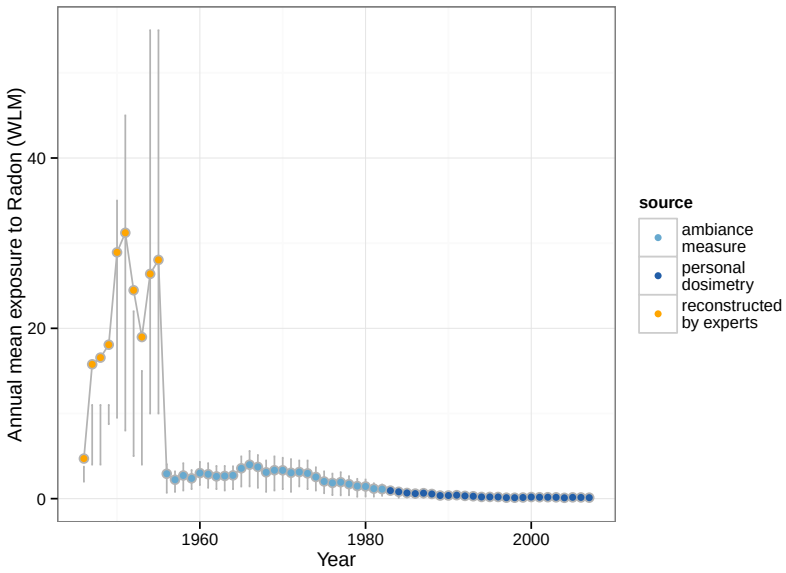

Figure 2.5: Annual mean radon exposure of exposed miners with 0.25 and 0.75 quantiles in the French cohort of uranium miners

Retrospective exposure reconstruction

In the early days of uranium mining, there was little awareness of the health risks that might be associated with the exposure to radon. Consequently, miners were not monitored for their radon exposure and all received exposure had to be reconstructed retrospectively [70, 33, 71] or extrapolated using exposure data collected in more recent years [72, 73]. In the French cohort of uranium miners, for instance, the exposure reconstruction for the years before 1956 was based on the knowledge of a group of experts, who were familiar with the exposure conditions during this period [74, 75]. In particular, the experts estimated the monthly exposure levels to radon for different mining sites based on the characteristics of the ore, the type of ventilation at the

different mining sites and a few measurements of radon concentration, that were available for the period before 1956. The monthly exposure estimates were then multiplied by the number of months a miner had worked at each mining site to derive individual annual exposure estimates [75].

Area sampling methods

Around 1950, indirect methods to measure the exposure to radon were developed and introduced in uranium mines in several countries for the purpose of radiation protection [46, 33, 72, 71, 68, 73]. In particular, these methods were introduced in 1949 in the Czech cohort of uranium miners [72], in 1955 in the Wismut cohort [71], in 1956 in the French cohort of uranium miners (see Figure 2.5) and in 1958 in the Ontario uranium miners cohort [73]. These indirect methods of exposure estimation were based on an area sampling strategy, where a certain number of measurements of either radon gas or radon progeny were taken in several areas of the mines [33, 73]. In the Wismut cohort, these ambient measurements were used to derive a job exposure matrix[6] [71]. In the other cohorts, they were generally used to derive individual exposure estimates by multiplying the time a miner had spent in a certain area of a mine by the exposure that was estimated in this area. Annual estimates for cumulated radon exposure were obtained by summing these terms over all areas a miner had worked during the year [68]. The number of measurements taken per year typically increased over time, leading to more and more precise estimates of radon exposure [72, 33, 73]. At the same time, other measures of radiation protection were introduced in uranium mines in the 1950s. In particular, it was decided in many countries to increase the ventilation or air flow in the mines in order to reduce the exposure levels, leading to a sharp decrease in the exposure to radon progeny in the early 1950s [46, 33, 77] for some cohorts and in the late 1960s [73] for others. The introduction of forced ventilation in the French uranium mines explains the sharp decrease in the average annual exposure to radon progeny, which can be observed in Figure 2.5.

Personal dosimetry

A program to develop personal dosimeters to measure radiation exposure in uranium mines was initiated in France in 1974 by the CEA [78]. These individual dosimeters, which permit to derive precise individual exposure estimates [69], were worn on the belt of a worker and allowed to simultaneously measure the worker's exposure to radon progeny, to long-lived radionuclides and to external gamma radiation. They were introduced in French uranium mines in 1983 (see Figure 2.5) and in the last operating mine in the Czech Republic in 2000 [69].

2.2.4 Results on the association between observed radon exposure and lung cancer mortality in cohort of uranium miners

[6] Job exposure matrices are used to estimate exposure to chemical and physical agents in occupational cohorts in the absence of individual measurements [76]. When developing a job exposure matrix, exposure levels are estimated for different job titles based on the tasks that are typically performed [15].

Current approach and risk estimates

Although the excess in lung cancer mortality in miners has been known for a long time [40, 41], the first systematic epidemiological studies to investigate the association between radon exposure and lung cancer in cohorts of uranium miners were only initiated in the early 1960s [79, 80]. Traditionally, when conducting statistical inference on the association between radon exposure and lung cancer mortality in cohorts of uranium miners, the most common approach is to adopt a grouped Poisson regression model. In this vein, the data are stratified into homogeneous groups of person-time according to variables like calendar period, attained age, duration of employment and cumulated radon exposure [81]. The number of lung cancer deaths that occurred in each stratum can then be modelled by a Poisson distribution with a rate parameter that is a function of the average cumulated radon exposure in the stratum [82]. In this context, the expected lung cancer mortality rate λ in a homogeneous category of person-years is defined as $\lambda = \lambda_0 \cdot (1 + ERR(X^{\text{cum}}))$, where λ_0 is the expected "baseline" lung cancer mortality rate and the Excess Relative Risk (ERR)[7] is modelled as a linear function of cumulated radon exposure X^{cum} (typically expressed in WLM): $ERR(X^{\text{cum}}) = \beta X^{\text{cum}}$ [83, 66]. β is the increase in ERR of lung cancer mortality that is associated with an increase in cumulated radon exposure of 1 WLM. Table 2.3 gives information on the association between radon exposure and lung cancer mortality in the most important cohorts of uranium miners. While studies conducted on these cohorts consistently show a significant association between radon exposure and lung cancer mortality, it is interesting to note that there is a substantial variability in the risk estimates for the different cohorts. In particular, the exclusion of all miners which were exposed during exposure periods characterised by retrospective exposure estimation leads to a marked increase in risk estimates in the French, the Czech and the Wismut cohort. In the French cohort of uranium miners, the exclusion of all miners exposed in the period before 1956 leads to the so-called "post-55" cohort [88, 81]. For the Ontario and the Eldorado cohort, estimates concerning lung cancer incidence are available, but they are very similar to the estimates for lung cancer mortality (1.30 [1.23;1.37] per 100 WLM for the Ontario cohort and 0.55 [0.37;0.81] per 100 WLM for the Eldorado cohort). In view of the low 5-year survival rate of patients suffering from lung cancer, it is not surprising that the analysis of lung cancer incidence and the analysis of lung cancer mortality result in virtually the same risk estimates [68].

Effect modifying variables

While the association between lung cancer mortality and radon exposure is classically described by a simple linear excess relative risk model, there is evidence that a number of variables modify this association [89]. In particular, the most important effect modifying variables of the association between radon and lung cancer include time since exposure [90, 47], attained age [47], age at exposure [72, 91], smoking [62] and exposure rate [70].

Concerning time since exposure, it was found in most cohorts of uranium miners that recent exposures (i.e., received in the last 15 years) were associated with higher risk estimates than exposures that were received long ago [80]. Analyses on sub-cohorts for which the smoking his-

[7] In other domains of epidemiology, it is common to model the Relative Risk (RR) via Poisson regression. The link between RR and ERR is $RR = 1 + ERR$. The parameters in an ERR model are constrained by the relation $ERR > -1$.

Table 2.3: Main characteristics and Excess Relative Risk (ERR) estimates for lung cancer mortality per 100 WLM of radon exposure obtained on the six most important cohorts of uranium miners

Study (Follow-up)	Number of miners	Number of deceased miners	Number (percentage) of miners deceased by lung cancer	Average exposure[a] (Min, Max)	ERR per 100 WLM [95% CI[b]]
Czech cohort [80] (31 Dec 2010)	9,978	5,286	1,141 (21.6%)	73 WLM	0.97 [0.74;1.27][c]
Post-68 cohort [80] (31 Dec 2010)	5,626	1,637	214 (13.1%)	7.5 WLM	2.0 [-0.1;5.3][c]
French cohort [81] (31 Dec 2007)	5,086	1,935	211 (10.9%)	36.6 WLM (0.01-960.1)	0.71 [0.31;1.30]
Post-55 cohort [81] (31 Dec 2007)	3,377	871	94 (10.8%)	17.8 WLM (0.01-128.4)	2.42 [0.90;5.14]
Colorado Plateau [63] (31 Dec 1990[d])	3,347		334	578.6 WLM	0.42 [0.30;0.70]
Eldorado cohort [68] (31 Dec 1999)	16,234		618	117 WLM	0.55 [0.37;0.78]
Ontario cohort [73] (31 Dec 2007)	28,546	8,318	1230 (14.8%)	21.0 WLM (0.0-875.1)	1.34 [1.27;1.42]
Wismut [86] (31 Dec 2003[d])	58,987	20,920	3,016 (13.8%)	280 WLM (0-3224)	0.19[e] [0.16;0.21]
Post-60 cohort [87] (31 Dec 2008)	26,766	3,820	334 (8.7%)	17 WLM (0-334)	1.3 [0.7;2.1]

[a] Average cumulated radon exposure, calculated on exposed miners

[b] 95% confidence interval

[c] 90% confidence interval.

[d] There are more recent follow-ups for these cohorts ([84] for the Colorado plateau cohort and [85] for the Wismut cohort), but we did not find published ERR estimates for these updated cohorts.

[e] When accounting for effect modification by age at median exposure, time since median exposure and exposure-rate, the central estimate is 1.08 per 100 WLM with a confidence interval of [0.69; 1.47].

tory of workers is available have observed a significant association between radon exposure and lung cancer mortality for both smokers and for non-smokers with an estimated risk coefficient that tends to be higher for non-smokers than for smokers [92, 47, 91]. Furthermore, case-control studies nested in the French cohort of uranium miners [93] and in two other European cohorts of uranium miners [94], found that, when adjusting for smoking, the effect of radon exposure on lung cancer risk persisted and the adjustment did not substantially alter the estimated risk

coefficient associated with radon exposure. Finally, the so-called inverse exposure rate effect describes the phenomenon that, for a given total cumulated exposure, the risk estimated for a worker is higher when this cumulated exposure is received over a long period of time than when it is received over a shorter period of time. While this effect is observed in most cohorts of uranium miners [66], it is not observed when restricting analyses to low levels of radon exposure [72, 71, 47, 80] or to subsamples for which precise exposure estimates are available [95, 70]. As the exposure levels generally decreased over time in cohorts of uranium miners while exposure estimates became more and more precise, it has repeatedly been suggested that the inverse exposure-rate effect may at least partly be due to exposure measurement error or exposure misclassification [96, 66, 97, 80]. In line with this hypothesis, Stram et al. (1999) found that the inverse exposure-rate effect observed in the Colorado plateau uranium miners cohort was weakened after the correction for measurement error [25].

2.3 The estimation of absorbed lung doses from radon progeny

While it is in theory possible to obtain precise radon exposure estimates, it is important to note that the potential health effects of this exposure heavily depend on the exposure conditions, for example on the type of physical activity an individual was engaged in during the exposure to radon. In particular, concerning the association between radon exposure and lung cancer mortality, it is more important how much energy is imparted to sensitive cells in the lung, which can be expressed via the concept of radiation dose, rather than how many decays occur outside of the body (which is the information that we may obtain by a radiation exposure value). In order to estimate the energy imparted to lung tissue, we need to model the deposition and the clearance of radioactive material in the lung, i.e. the mechanism by which radioisotopes deposit in the lung and by which they are transported to other organs. This section will give a brief overview of the concept of radiation dose and the estimation of absorbed lung dose from radon progeny.

2.3.1 The concept of radiation dose

Ionising radiation can give rise to a variety of adverse health effects on living organisms. The traditional unit to measure radioactivity is named Becquerel after Henri Becquerel. However, this unit, which corresponds to the expected number of decays of a radioactive material per second [50], is not always suitable to reflect the potential health effects of radiation exposure. Indeed, the harm caused by ionising radiation may depend on a number of factors, including exposure conditions, the radiosensitivity of cells within the organism and the type of ionising radiation [98]. In order to account for these differences, three different dose quantities are commonly used in radiation protection to reflect the biological effects of ionising radiation, namely organ absorbed dose, equivalent dose and effective dose.

Organ absorbed dose

Organ absorbed dose is a quantity representing the mean energy per unit mass imparted to tissue being irradiated by ionising radiation [99, 29]. The unit of absorbed dose is gray (Gy) with one gray being equivalent to 1 joule per kilogram.

Equivalent dose

While the unit of organ absorbed dose is more suitable to reflect the biological effects of different radioisotopes than Becquerel, one can only achieve a limited comparability of different types of radiation with this dose quantity. Indeed, different types of radiation vary in the way that they interact with matter. Therefore, a given energy deposited by α- and γ-radiation, for instance, will not have the same biological effectiveness, since α-particles are more heavily charged and deposit their energy more densely. In order to derive a dose quantity that allows to compare different types of radiation, absorbed dose has to be multiplied by a so-called radiation weighting factor. Equivalent doses are classically expressed in sievert (Sv) [100].

Effective dose

Finally, different biological tissues can be more or less sensitive to ionising radiation. Therefore, it is indispensable to derive a dose quantity that can account for this radiosensitivity in order to be able to compare the health risks caused by different radioisotopes. This dose quantity is called effective dose and it accounts for both the biological effectiveness of radiation and the radiosensitivity of the organ or tissue being irradiated [100].

2.3.2 The human respiratory tract model

Absorbed lung doses from radon cannot be observed directly but have to be modelled using mathematical models [101]. The estimation of absorbed lung dose due to the inhalation of radon is commonly based on the so-called human respiratory tract model [102], which is illustrated in Figure 2.6. This model is used to describe the deposition and the clearance of radioisotopes in the respiratory tract in order to derive absorbed radiation doses delivered to different regions of the lung. It divides the human respiratory tract into the extrathoracic airways and the thoracic airways. The latter can be divided into the bronchial region (BB), the bronchiolar region (bb) and the alveolar-interstitical (AI) region [104].
While radon can also give rise to doses absorbed to other organs, absorbed lung doses will dominate these doses, as most of the inhaled radon progeny will decay in the lung before clearance to other organs can take place.

2.3.3 Uncertain input parameters in dose estimation

As in many occupational and environmental settings involving dose estimation [105, 106, 107], the estimation of absorbed organ doses in uranium miners is subject to many sources of uncertainty. A number of studies have been conducted to identify the most influential input parameters in absorbed lung dose calculation for radon progeny [108, 37, 38, 109]. The input parameters which are most likely to influence the estimation of absorbed lung dose include the unattached fraction [37], breathing rate [108, 101, 37, 60, 110], the target cell depth [37] and the activity size distribution of attached and unattached radon progeny [101, 60, 111, 104]. As these input parameters are in general not measured, they are highly uncertain, thereby introducing uncertainty in the calculation of absorbed lung dose.

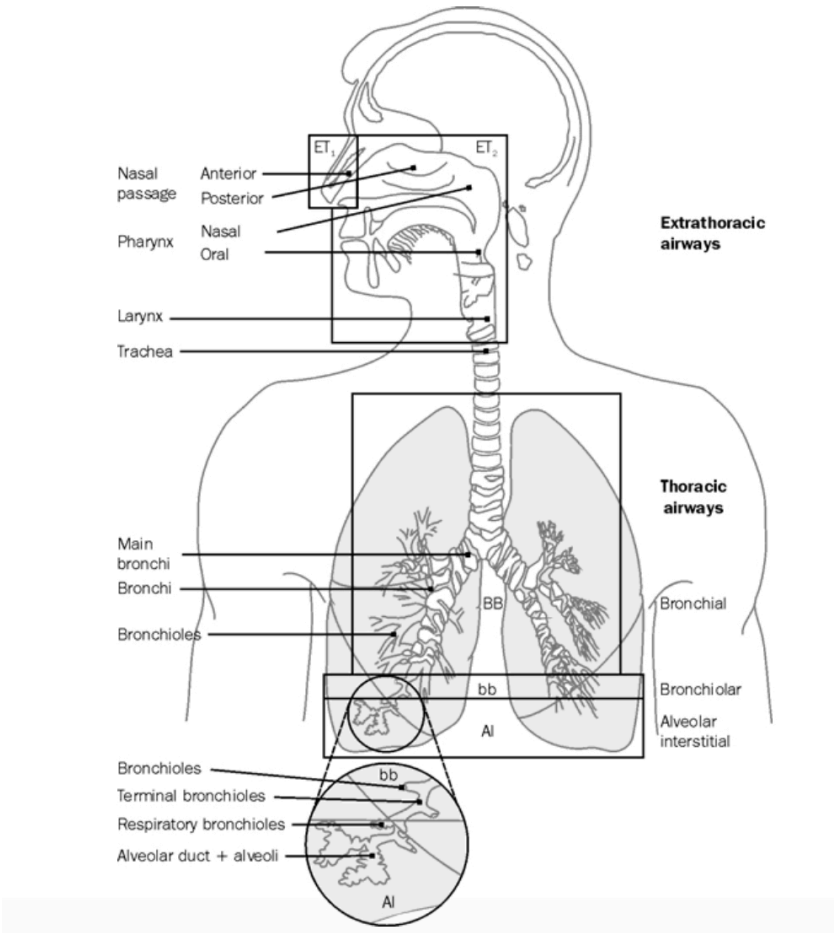

Figure 2.6: The human respiratory tract model dividing the lung into the extrathoracic airways (ET_1 and ET_2), the bronchial region (BB), the bronchiolar region (bb) and the alveolar-interstitial regions (AI). Source: [103]

Unattached fraction

As mentioned in section 2.2.1, when inhaled, almost all radon gas will subsequently be exhaled [59]. At the same time, radon decays into a series of short-lived radioisotopes, which have higher chemical reactivity and can therefore react rapidly with trace gases and vapours and form clusters of around 1 nanometer [60]. Commonly, these particles are referred to as the unattached fraction of radon progeny. Unattached particles can attach to aerosol particles, which are present in the mine atmosphere. In this case, they are referred to as the attached fraction of radon progeny [60]. Attached and unattached radon progeny can deposit in the respiratory tract [112], but the deposition probability and the regions in which the particles will deposit are very different for attached and unattached radon progeny. The deposited progeny will either decay in the lung or be transferred to the gastrointestinal tract or to other organs via clearance mechanisms [113], but due to their relatively short half-lives, most of the radon progeny decays in the lung before clearance can take place [59, 88, 114, 40].

Average breathing rate

Breathing rate is an important input parameter in the estimation of absorbed lung dose, as the number of radioisotopes inhaled by a miner can be assumed to be directly proportional to his breathing rate [39, 60]. At the same time, it is very rare to dispose of direct and individual measurements of breathing rate in underground miners. Several approaches exist to estimate the average breathing rate of an individual in the absence of direct measurements. As the average breathing rate depends on the level of physical activity, it is common to estimate the fraction of time a miner spent in different activities and to multiply the determined proportions of time by a fixed value for breathing rate that is assumed to be known for this type of activity [109, 40, 60].

Target cell depth

A large part of lung dose is delivered by α-particles [88], in particular due to the disintegration of ^{218}Po and ^{214}Po [61] . As α-particles have a very short range, travelling only some millimetres in tissue, the depth of target cells is an important parameter in dose calculation. The fraction of alpha particle energy deposited in the target cells, and hence the absorbed dose, will depend on their depth and on the thickness of different layers that separate these cells from the interior of the human lungs [40].

Activity size distribution

The fraction of radon progeny that are attached to ambient aerosol particles and the activity size distribution of the attached and the unattached radon progeny play an important role in the estimation of absorbed dose [115]. Indeed, the deposition of an inhaled particle in the lung strongly depends on the particle size [37, 111, 40]. The activity size distribution depends on the exposure conditions in the mines, in particular the presence of diesel engines [116, 60].

Summary

In conclusion, when analysing the association between the exposure to radon and its progeny and lung cancer mortality in cohorts of uranium miners, one is faced with a certain number of uncertainties. The main sources of uncertainty, which are summarised in Figure 2.7, do not only include exposure measurement error and other uncertain input parameters in dose calculation, but also the potential for a modification of the exposure-response relationship by time since exposure, smoking and the exposure rate effect. If our aim is to account for these sources of uncertainty, we are faced with an additional source of uncertainty, namely our uncertainty on the models to describe the relationship between the different quantities shown in Figure 2.7. There were substantial changes in the methods of exposure assessment and in the exposure conditions over time in almost all the cohorts of uranium miners. It has been suggested some of the effect modifying variables, in particular exposure rate, may at least partially be a result of measurement error [96, 66, 97, 80]. In other words, if we do not account for the uncertainty

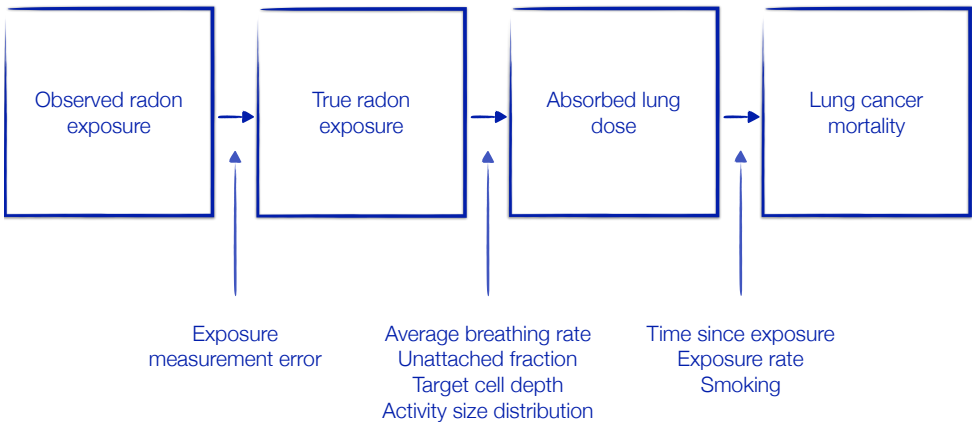

Figure 2.7: Main sources of uncertainty in the analysis of the association between radon exposure and lung cancer mortality in cohorts of uranium miners. Modified from BEIR VI [63].

in the association between observed radon exposure and absorbed lung dose (on the left-hand side in Figure 2.7), these uncertainties may create apparent distortions of the dose-response relationship (on the right-hand side in Figure 2.7).

Chapter 3

Measurement error in epidemiological studies

The aim of most epidemiological studies is to evaluate the association between a disease outcome and a number of risk factors or exposure variables, which are not all known perfectly. Indeed, epidemiology is primarily an observational science where the exposure conditions of study participants are not under experimental control [3, 117], contrary to many other disciplines. In this situation, the determination of the true exposure of study participant i, X_i, may be difficult, expensive or even impossible and it is common to use questionnaires, interviews or a strategy of group-level exposure assessment in order to obtain an error-prone surrogate Z_i for true exposure X_i. The association between the surrogate, or observed exposure, Z_i and the disease outcome Y_i, described by a so-called disease model, is then studied to infer the association between true exposure X_i and Y_i, resulting in a measurement error problem. While the term "measurement error" is traditionally reserved to describe a lack of precision in a measuring device in the physical sciences, it has a very broad definition in epidemiological studies [118, 119]. Measurement error of study participant i, denoted U_i hereafter, refers to the discrepancy between the observed exposure value Z_i, i.e., the exposure value assumed in an epidemiological study, and the true (and unknown) exposure value X_i, regardless of the reasons for this discrepancy. Another specificity of measurement error in epidemiological studies arises from the fact that the exposure to a certain chemical or physical agent is typically not constant over time [120, 14]. Therefore, it may be more adequate to describe the discrepancy of true $X_i(t)$ and observed exposure $Z_i(t)$ of study participant i at time t by the measurement error term $U_i(t)$, depending on time t.

The aim of this chapter is to give an overview of the literature on the different types of measurement error, their potential effects on risk estimates in epidemiological studies and of standard approaches to correct for these effects. Measurement error has attracted considerable attention in the epidemiological and in the statistical literature. While problems related to exposure measurement error occur in all areas of epidemiology, the study of the effects of measurement error in risk estimation and the development of methods to account for it have received the greatest attention in nutritional epidemiology [121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134], radiation epidemiology [27, 26, 28, 135, 29, 136, 137, 138, 107, 32, 35, 139] and the study of the health effects of air pollution [140, 141, 142].

The popularity of this topic in radiation epidemiology might be linked to the fact that the main

risk factor of interest in many studies in this domain is radiation dose (see section 2.3.1) [137]. Radiation dose is an abstract concept to express the biological effects of radiation exposure. The dose absorbed to critical organs, for instance cannot be measured directly, but has to be calculated by combining physical or biological measurements with mathematical models involving a number of uncertain input parameters which have to be assumed to be known for this purpose [143, 118]. In this chapter, we will adopt a broad definition of the term measurement error to describe both exposure and dose uncertainty in epidemiological studies. Moreover, we will limit our discussion to cases where the exposure variable is prone to error, although measurement error can also occur in the disease outcome.

3.1 Measurement error and its consequences on inference

It is widely acknowledged that exposure measurement error can lead to misleading results in observational studies when it is not or only poorly accounted for, potentially causing bias in risk estimates, a reduction in statistical power and a distortion of the exposure-response relationship [105, 8, 144]. However, the term measurement error is extremely broad [118, 119] and its exact consequences on statistical inference depend both on the type of error and on the type of disease model.

As can be seen in the following, many important properties of measurement error can be resumed by independence assumptions. In particular, in the case of independence between measurement error of study participant i at time t, $U_i(t)$, and the corresponding value of true exposure $X_i(t)$, we will call this error term classical measurement error. On the other hand, if $U_i(t)$ is independent of observed exposure $Z_i(t)$, the error term is called Berkson error. The condition of non-differential measurement error is satisfied if the error term $U_i(t)$ is independent of the outcome Y_i. Finally, if both the error terms of different participants at a certain time t are independent (i.e. $U_i(t) \perp\!\!\!\perp U_{i'}(t)\ \forall i \neq i'$) and the error terms of a certain participant i are independent over time, (i.e. $U_i(t) \perp\!\!\!\perp U_i(t')\ \forall t \neq t'$), then the error term is said to be unshared.

3.1.1 Different types of measurement error

Multiplicative and additive measurement error

Depending on whether one assumes an additive or a multiplicative measurement model, the error term $U_i(t)$ will either describe the difference or the ratio between true $X_i(t)$ and observed exposure $Z_i(t)$. While it is common to assume an additive model and a normal distribution to describe the variability of $U_i(t)$ in the statistical literature [145, 8, 146], this model may be problematic in epidemiological studies as both true and observed exposure typically take non-negative values. Moreover, the empirical distribution of observed exposure [31] and the nature of error sources may often suggest a multiplicative model and a log-normal distribution of $U_i(t)$ [147, 148, 137, 149, 150, 151, 152, 29, 153]. Consequently, we will in general adopt a multiplicative error model in the following to account for the nature of exposure measurement error in epidemiological studies, but this work can easily be extended to the case of additive measurement error.

Berkson and classical measurement error

The most common distinction that is made in the literature to classify measurement error is the difference between Berkson error and classical measurement error [8, 106, 10].
In the classical measurement model, we assume that true exposure $X_i(t)$ of study participant i at time t is observed via its error-prone surrogate $Z_i(t)$. $Z_i(t)$ is considered to be a random variable that can be modelled as a function of true and unknown exposure $X_i(t)$ and an error term $U_i(t)$ that is independent of true exposure $X_i(t)$. For a study participant i at time t, the multiplicative classical measurement model is given by:

$$Z_i(t) = X_i(t) \cdot U_i(t). \tag{3.1}$$

Classical error arises naturally in cases where exposure is measured individually for each study participant via a measurement device. Thereby, the classical measurement model is used to describe the fact that if we used several devices to measure the exposure of a given study participant at the same time, the observed values on different measuring devices would usually indicate slightly different values [105].
Berkson error, on the other hand, arises in cases where the true exposure values $X_i(t)$ of study participants deviates from a fixed and known exposure value $Z_i(t)$:

$$X_i(t) = Z_i(t) \cdot U_i(t), \tag{3.2}$$

where the measurement error term $U_i(t)$ is independent of $Z_i(t)$. Originally proposed by Berkson (1950) [154], this model is primarily intended to describe the error occurring in an experimental setting with a predefined exposure value $Z_i(t)$ where the true exposure value $X_i(t)$ deviates from this prefixed value due to imprecisions in the realisation of the experiment [155, 156]. For instance, one can imagine a X-ray machine, which is programmed to deliver a certain radiation dose in medical imaging. The actual dose received by a patient will slightly vary depending on the exposure conditions, for instance as a function of his size or of his exact position with respect to the machine.
In epidemiological studies, this model is also commonly assumed in situations, where a group-based strategy is used to determine individual exposure [14, 147, 31, 32, 29, 106, 119]. The idea behind this assumption is that, in the case of a group-based strategy, we observe an exposure value $Z_j(t)$ which is common to all participants i in a group j. The individual exposure value of participant i in group j at time t, $X_{ij}(t)$, will in general randomly deviate from this observed exposure value $Z_j(t)$. This view has been repeatedly criticised, however, because it suggests that the value measured at the group-level is known without error [26, 105, 157]. In order to obtain a more realistic model to describe measurement error occurring in a group-based exposure assessment strategy, it has been suggested to use a mixture of Berkson and classical measurement error [26, 10, 158], where the classical component is intended to reflect the error in the measurement process of the group-based exposure value, while the Berkson component describes the deviation of the individual exposure values from this group-based exposure value.

Differential and non-differential measurement error

Another fundamental distinction that can be made to describe measurement error in epidemiological studies is the difference between differential and non-differential measurement error.

Measurement error is called non-differential when the error term is independent of disease status [159]. In other words, the assumption of non-differential error describes the conditional independence of observed exposure $Z_i(t)$ and disease outcome Y_i given true exposure $X_i(t)$. The assumption of non-differential measurement error seems reasonable when exposure is assessed in a prospective fashion, or in other words before the occurrence of a disease. In case-control studies, where the values on exposure variables are obtained after the diagnosis of the disease, differential measurement error may arise because of recall bias, i.e. cases and controls do not remember their exposure histories in the same way [160, 161, 127, 162]. In particular, it is possible that cases spent more time trying to remember their past exposures, because they are trying to find explanations for their present disease and this could lead to smaller measurement errors for cases than for controls. On the other hand, it is also possible that cases overestimate their past exposure while controls could have a tendency to underestimate their exposure.

Systematic and random error

When treating measurement error, it is common to assume that there is only a random error component $U_i(t)$, i.e. $Z_i(t) = X_i(t) \cdot U_i(t)$ with $E(Z_i(t)|X_i(t)) = X_i(t)$ for classical measurement error or $X_i(t) = Z_i(t) \cdot U_i(t)$ with $E(X_i(t)|Z_i(t)) = Z_i(t)$. Apart from this random error component, we can also assume that there is a systematic bias in the discrepancy between true $X_i(t)$ and observed exposure $Z_i(t)$. For instance, we can assume in a classical measurement model that we are not only faced with the random classical measurement error component $U_i(t)$, but also with a bias term a, i.e. :

$$\begin{aligned} Z_i(t) &= a \cdot X_i(t) \cdot U_i(t) \text{ with} \\ E(Z_i(t)|X_i(t)) &= a \cdot X_i(t). \end{aligned} \tag{3.3}$$

This bias affects the exposure measurements of all study participants in the same way and will lead to a systematic over- or underestimation of true exposure for the whole sample [163, 164]. Systematic measurement error may arise because of errors in the calibration of a measurement device which is used for all study participants [164].

Heteroscedastic measurement error

In epidemiological studies, exposure values for different study populations may be measured with varying degrees of precision. In particular, in occupational cohort studies, it is common to use different methods of exposure assessment for different exposure periods. In this case, exposure uncertainty can be adequately modelled by a heteroscedastic measurement model to describe differences in the error variances for different exposure periods. More generally, we can assume a heteroscedastic measurement model whenever the error variance depends on a covariate $V_i(t)$. For instance, in a multi centre study taking place in several hospitals, we can imagine that the machines and the procedures for the measuring of the explanatory variables in the centres differ. In this case, we can describe the resulting differences in the magnitude of measurement error for the patients in these centres by a heteroscedastic error structure.

Shared measurement error

Recently, there has been growing interest in measurement error components, which are shared for several individuals in radiation epidemiology [165, 105, 106, 118, 119, 138, 107, 153]. These shared measurement errors arise when the discrepancy between true and observed exposure depends on an error component that affects the exposure or dose estimates for several study participants in the same way. These error components may often arise in the retrospective reconstruction of radiation doses received by a study population. Indeed, a retrospective dose reconstruction typically involves many unknown quantities, which have to be estimated. For instance, in the reconstruction of thyroid dose estimates for a cohort of Belarusian children exposed due to the Chernobyl accident, mathematical models are used to reconstruct the transport of iodine-131 from the ground to a child's thyroid [107]. In this situation, input parameters like the daily deposition of iodine-131 in a given settlement or the daily grass consumption of a cow have to be estimated to derive thyroid doses. Errors in the estimation of the first parameter will affect all children of a given settlement in the same way. According to Drozdovitch et al. (2015), the errors occurring in the second parameter will even affect all study participants in the cohort simultaneously [107].

While researchers in radiation epidemiology are mainly interested in measurement errors that are shared between individuals, there has long been awareness of measurement error components that are shared for several exposure values of the same individual in nutritional epidemiology [123, 124, 125, 166, 132]. For instance, if a study participant has a tendency to underestimate his calorie intake in a questionnaire, it is likely that he will show this tendency every time he completes the questionnaire, thereby creating a subject-specific error component. Likewise, error components that are shared for several years of exposure may arise in occupational cohort studies when a method of group-based exposure assessment is used, for instance when assessing exposure via job-exposure matrices. As described in section 2.2.3, when developing a job-exposure matrix, exposure levels are estimated for different job titles based on the tasks that are typically performed [15] and all workers with a given job title are affected the same exposure value. However, differences in working conditions, individual worker practices and other worker characteristics may lead to varying exposure levels within a given job category. As these working conditions and worker characteristics typically remain constant for several exposure measurements [16, 15], this may lead to subject-specific error components. For instance, workers in a car manufacturing plant have been reported to receive very different long term average exposures to isopropyl alcohol as a result of their differences in body length, although they performed basically the same tasks [16].

In summary, shared measurement errors can either describe a statistical dependence of errors occurring for different study participants or a statistical dependence of errors of the same study participant at different time points. In the following, we will consider that an error component is shared for all years of exposure of the same individual if $U_i(t) = U_i \; \forall t$. A Berkson error shared for several years of an individual can thereby be expressed in the following way:

$$X_i(t) = Z_i(t) \cdot U_i. \tag{3.4}$$

Similarly, an error component shared for all individuals in a study can be expressed by $U_i(t) = U(t)\ \forall i$. We could express a classical measurement error shared between individuals by:

$$Z_i(t) = X_i(t) \cdot U(t). \tag{3.5}$$

However, in general, it is more convenient to assume that an error component shared within individuals is shared for a certain sub-group rather than for all individuals in a study. In the same way, we can assume that an error component shared within individuals, i.e. for several years of the same individual, is only shared for the exposure years in a specific exposure period, rather than for all exposure years of an individual.

3.1.2 The effects of measurement error on statistical inference

When effect size is tiny and measurement error is huge, you're essentially trying to use a bathroom scale to weigh a feather - and the feather is resting loosely in the pouch of a kangaroo that is vigorously jumping up and down. - Andrew Gelman in a post on his blog (21 April 2015)

Despite the variety of types of measurement error that can occur in epidemiology, studies that assess the effects of measurement error on statistical inference and methods to correct for measurement error mainly focus on errors that are homoscedastic, non-differential and unshared [153, 167]. Moreover, many of the assumptions that are generally made concerning the effects of measurement error on statistical inference are shaped by analytical results that are available for the simple linear regression model with only one covariate that is prone to additive error. This model, which describes the conditional expectation of a continuous outcome variable as a function of an error-prone covariate, has received great attention in the measurement error literature [145, 168, 169, 26, 128, 8]. However, in epidemiology, as the study of the patterns and the causes of disease, the relevance of this model is extremely limited since it is only rarely possible to model disease outcomes as uncensored continuous variables [170]. Disease outcomes

are more commonly described by binary variables (e.g. presence or absence of disease), failure times (e.g. time until the disease is declared) or count variables (e.g. number of cases of a certain decease). Consequently, the effects of measurement error in logistic regression, proportional hazards and Poisson regression models are of more interest in epidemiological studies [171, 137]. In this section, we will give a brief overview of the analytical results available in linear regression and then focus on logistic regression and proportional hazards models. Finally, we will briefly mention some results concerning the effects of measurement error in Poisson regression and the possibility of measurement error to lead to an attenuation of the exposure response relationship in occupational cohort studies.

Measurement error in simple linear regression

Consider the simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \tag{3.6}$$

where the continuous outcome of study participant i, Y_i, is modelled as a function of the value he takes on the explanatory variable X_i. β_0 and β_1 are unknown regression coefficients and the error terms ϵ_i are independently and identically distributed with $E(\epsilon_i) = 0$ and variance σ_ϵ^2.
Suppose X_i can merely be observed through Z_i which is a surrogate for X_i that is prone to non-differential and additive classical measurement error, i.e. $Z_i = X_i + U_i$, implying that the error term U_i is independent of both X_i and Y_i. We will further assume that the error terms U_i are independently and identically distributed for different study participants and denote the variance of U_i as σ_u^2 and the variance of X_i as σ_x^2. The standard approach to estimate β_1 consists in neglecting measurement error by fitting the model:

$$Y_i = \beta_0^* + \beta_1^* Z_i + \epsilon_i^* \tag{3.7}$$

where X_i is replaced by its surrogate Z_i. In this model, the ordinary least squares estimate of β_1 is given by $\hat{\beta}_1^* = \frac{Cov(Z_i,Y_i)}{Var(Z_i)}$. As U_i is independent of X_i, the denominator takes the value $Var(Z_i) = Var(X_i) + Var(U_i) = \sigma_x^2 + \sigma_u^2$. Moreover, $Cov(Z_i, Y_i) = Cov(X_i, Y_i) + Cov(U_i, Y_i)$. Due to the assumption of non-differential measurement error the second term is zero, resulting in $\hat{\beta}_1^* = \frac{Cov(X_i,Y_i)}{\sigma_x^2+\sigma_u^2}$. When comparing this expression with the ordinary least square estimate of β_1 given by $\hat{\beta}_1 = \frac{Cov(X_i,Y_i)}{\sigma_x^2}$, which is known to be unbiased in cases where X_i is observable, we remark that the estimator $\hat{\beta}_1^*$ is biased towards the null by an attenuation factor λ, i.e. $\hat{\beta}_1^* = \lambda\hat{\beta}_1$ where $\lambda = \frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}$.
Additive and non-differential Berkson error, on the other hand, does not bias risk estimates in the simple linear regression model [8, 160]. In this case, we have $X_i = Z_i + U_i$ with U_i independent of Z_i and $E(U_i|Z_i) = E(U_i) = 0$. We can write

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i && (3.8)\\ &= \beta_0 + \beta_1(Z_i + U_i) + \epsilon_i && (3.9)\\ &= \beta_0 + \beta_1 Z_i + \beta_1 U_i + \epsilon_i && (3.10)\\ &= \beta_0 + \beta_1 Z_i + \epsilon_i^*. && (3.11) \end{aligned}$$

The new error term $\epsilon_i^* = \beta_1 U_i + \epsilon_i$ is centred at zero since $E(\epsilon_i^*) = E(\beta_1 U_i + \epsilon_i) = \beta_1 E(U_i) + E(\epsilon_i) = \beta_1 \cdot 0 + 0 = 0$. Under the hypothesis that U_i and ϵ_i are independent, the variance of ϵ_i^*

can be expressed as $Var(\beta_1 U_i + \epsilon_i) = Var(\beta_1 U_i) + Var(\epsilon_i) = \beta_1^2 \sigma_U^2 + \sigma_\epsilon^2$. Therefore, estimating β_1 in 3.11 or in 3.8, for instance via ordinary least squares, will yield an unbiased estimate of β_1. However, as $\beta_1^2 \sigma_U^2 + \sigma_\epsilon^2 > \sigma_\epsilon^2$ for $\beta_1 \neq 0$ and σ_U^2, Berkson error in the simple linear regression model may lead to a widening of confidence intervals and thereby to a loss in statistical power. In summary, when there is only one error-prone covariate in the simple linear model, non-differential classical measurement error will lead to an underestimation of the regression coefficient β_1, where the bias depends both on the variance of measurement error σ_u^2 and the variance of true exposure σ_x^2. In the case of Berkson error, however, the estimator $\hat{\beta}_1^*$ is unbiased. These results are regularly repeated in the discussion of results of epidemiological studies [172, 173] and in papers that give recommendations concerning measurement error in epidemiology [3, 174, 175, 147, 160, 105, 118, 119]. Unfortunately, they are only rarely accompanied with the warning that these statements are only valid in the simple linear regression model, which is hardly ever used in epidemiological studies and under the condition that only one covariate is prone to error. If more than one covariate is prone to measurement error, which is virtually always the case in epidemiological studies, it is more difficult to predict whether measurement error will attenuate or inflate risk estimates [122, 126, 129]. More specifically, in multiple linear regression, measurement error in one covariate can induce bias in the estimates of all regression coefficients, including the coefficients associated with covariates that are not measured with error [162]. In many epidemiological studies, the effects of measurement error will therefore rather resemble Andrew Gelman's kangaroo jumping vigorously up and down on a bathroom scale, while you are trying to weight a feather that is in its pouch, than the very predictable effects that are often assumed. Despite the efforts of many researchers to unveil this misunderstanding [167, 176, 177, 4, 178, 129], it seems as if there was still a wide-spread belief in epidemiology that Berkson error does not cause bias, while classical measurement error can only bias risk estimates towards the null [4, 12].

Measurement error in logistic regression

The logistic regression model describes the conditional probability of a binary outcome Y_i for study participant i given the values he takes on a covariate X_i as

$$P(Y_i = 1|X_i) = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}. \tag{3.12}$$

Contrary to the simple linear regression model, both Berkson and classical measurement error can lead to biased risk estimates in logistic regression [26, 28, 4]. There are no general formulas indicating the exact magnitude of bias that is introduced by measurement error in logistic regression [8]. Heid et al. (2002) [28] provide approximate expressions, showing that, in the case of classical additive error, the bias introduced in risk estimates in logistic regression depends on the measurement error variance and the variance of true exposure. Moreover, Burstyn et al. (2014) [179] perform a simulation study that demonstrates that non-differential exposure misclassification can not only attenuate risk estimates, but also lead to an increase in false positive findings in logistic regression, i.e., it increases the chance of erroneously observing a significant association between the exposure and the disease outcome of interest in cases where this association does not exist.

Measurement error in proportional hazards models

The outcome of many epidemiological studies can be best described as a failure time, e.g. as the time until the occurrence of a certain event (typically the occurrence of or the death by a certain disease). This type of outcome is commonly right-censored, i.e., we do not observe the failure time of interest T_i for all study participants. Instead, we observe $Y_i = \min(T_i, C_i)$, where C_i is the censoring time of study participant i and a censoring indicator δ_i: $\delta_i = 1$ if $T_i \leq C_i$ and $\delta_i = 0$ otherwise. The most popular choice to describe the association between an exposure and a right-censored failure time is the proportional hazards model. In its standard form, this model describes the instantaneous hazard rate of experiencing the event of interest for study participant i, $h_i(t)$, by:

$$h_i(t) = h_0(t)\varphi(X_i, \beta) \tag{3.13}$$

where $h_0(t)$ is the so-called baseline hazard, which usually corresponds to the instantaneous hazard rate for an unexposed study participant (i.e. $X_i = 0$). $\varphi(X_i, \beta)$ is a positive term representing the hazard ratio between a study participant who received an exposure of X_i and a non-exposed study participant. This model was first introduced by Cox (1972) [180], who proposed a log-linear model by specifying $\varphi(X_i, \beta) = \exp(X_i\beta)$. Moreover, he proposed a semi-parametric method of partial likelihood maximisation, which allows to maximise the likelihood with respect to the risk coefficient β, without having to specify the baseline hazard $h_0(t)$.
The most fundamental result concerning the effect of measurement error in the proportional hazard model is that this partial likelihood method can no longer yield valid inference in general in the presence of measurement error, because the baseline hazard and the effect of the exposure variables are no longer separable [181, 182, 183]. Moreover, similar to logistic regression, both Berkson and classical measurement error can lead to biased risk estimates in proportional hazards models [4, 32]. Hugh (1993) shows that in the case of a single explanatory variable which is observed with non-differential and additive classical measurement error that is normally distributed, the attenuation factor λ is approximately equal to the attenuation factor in linear regression (i.e. $\frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2}$) for cases when the disease is rare (i.e. the failure times of a high percentage of the sample are censored). Concerning the shape of the exposure-response association, Keogh et al. (2012) [184] conducted a simulation study and found that when the true association between an exposure and the outcome is nonlinear, unshared classical measurement error makes this association appear more linear.

Measurement error in Poisson regression

When analysing cohort data to evaluate the association between the exposure to a certain chemical or physical agent and a disease outcome, some advocate Poisson regression as the method of choice [185]. This approach, which is particularly popular in radiation epidemiology [139], typically entails the tabulation of person-time and events in order to model the number of disease cases in a certain sub-group or stratum. To achieve this tabulation, all explanatory variables have to be categorised [186]. On the first glance, the categorisation of an error-prone exposure variable may seem a simple and straightforward solution to alleviate the impact of measurement error. Tielemans (1998) [175], for instance, argues that the categorisation of error-prone exposure should lead to Berkson type measurement error and therefore result in almost unbi-

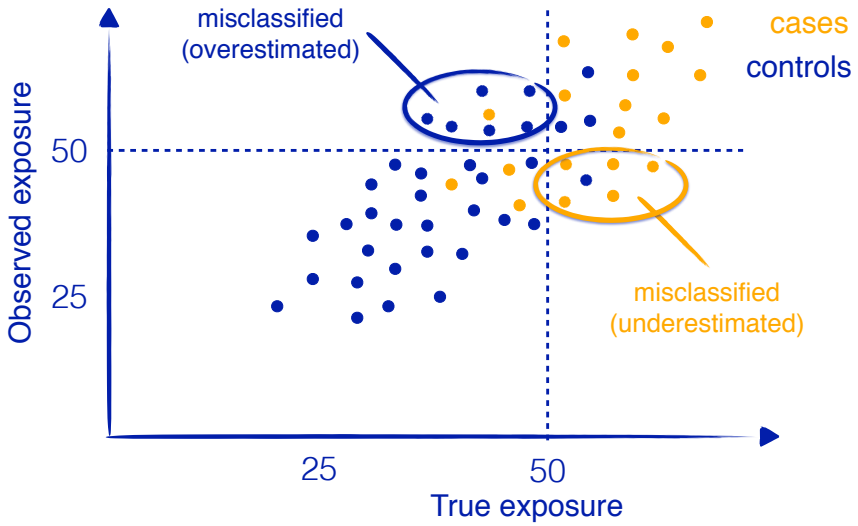

Figure 3.1: Illustration of the problems arising in the categorisation of a error-prone continuous exposure variable

ased risk estimates. As mentioned earlier, the assumption that Berkson type measurement error leads to unbiased risk estimates is only true in simple linear regression and this model is only very rarely suitable for the description of disease outcomes in epidemiological studies. More importantly, there is some evidence that the categorisation of error-prone exposure variables can actually aggravate the problem of exposure measurement error in epidemiological studies. If the exposure of interest is associated with the outcome, the true exposure distribution will not be the same for cases and for controls. As the probability of misclassification depends on true exposure (being higher around the category limits), the categorisation of a continuous exposure variable presenting non-differential measurement error is very likely to lead to differential misclassification [187, 82, 188]. Figure 3.1 illustrates this phenomenon. Although the probability of misclassification in this example is comparable for cases and for controls, there is a tendency to underestimate the exposure for cases and to overestimate the exposure of controls when the error prone exposure variable is categorized. Differential misclassification can create substantial bias in risk estimates [186] and, in this case, methods that assume non-differential misclassification can yield highly misleading results [187].

The potential of measurement error to lead to an attenuation in the exposure-response relationship

In occupational cohort studies, one frequently observes an attenuation of the exposure-response curve for high exposure values [17]. It has been suggested that this phenomenon could be explained by the fact that exposure uncertainty and the magnitude of exposure are both highest for the earliest exposure periods, which are often characterised by a method of retrospective exposure assessment [189, 17, 14]. Steenland et al. (2015) [14] recently examined this hypothesis on simulated data and merely found a modest attenuation at high exposure values. However, the authors treated cumulative exposure in an occupational cohort as time fixed variable known at baseline, thereby ignoring its time-varying nature and the possibility of exposure uncertainty

components to be shared for several workers or for several exposure values of the same worker.

3.2 Standard methods to correct for measurement error

In light of the severeness of the potential consequences of exposure measurement error on statistical inference, it seems surprising that measurement error is only rarely accounted for in epidemiological studies [12, 190, 191]. At the same time, there is a plethora of approaches that have been proposed in the statistical literature to account for non-differential measurement error. Examples range from simple methods, like regression calibration, which only allow for an approximate correction of parameter estimates in nonlinear regression models [192, 144], to elaborate methods, which can produce consistent parameter estimates in both linear and nonlinear models if the assumed models are specified correctly.

The starting point of basically all methods to account for measurement error is the specification of a response model and a measurement model. In epidemiology, the response model is also frequently referred to as disease model [193, 194, 160]. It relates disease outcome Y_i of study participant i to his true and unknown values of exposure $X_i(t)$ at time t. In the case of Berkson error, the measurement model describes the distribution of the true exposure of study participant i at time t, $X_i(t)$, conditional on the corresponding observed value $Z_i(t)$ (see equation 3.2). In the case of classical measurement error, on the other hand, the measurement model describes the distribution of the observed exposure of study participant i at time t, $Z_i(t)$, conditional on the corresponding true exposure value $X_i(t)$ (see equation 3.1). So-called functional methods only specify a disease and a measurement model when correcting for classical measurement error, while structural methods additionally require a parametric model to describe the probability distribution of the true and unknown exposure $X_i(t)$. Thereby, they assume that the distribution of the vector of exposure values X belongs to a known family of parametric distributions [195]. The specification of this so-called exposure model may involve some difficulty because it involves hypotheses about a latent variable, which by definition is not observed. In functional methods, on the other hand, the true exposure values $X_i(t)$ are either regarded as unknown and fixed or, if $X_i(t)$ is considered to be a random variable, only minimal assumptions about its distribution are made [8], for instance by modelling the expected values of $X_i(t)$ conditional on $Z_i(t)$. If the exposure model is specified correctly, structural methods have been shown to outperform functional methods in terms of efficiency [160, 196, 197, 198, 199]. This section will give a brief overview on the possible approaches by presenting regression calibration, simulation extrapolation, likelihood-based approaches and methods based on multiple realisations of exposure or dose estimates.

Alternative approaches, which are not treated in more detail here, include the instrumental variables approach [169, 198], multiple imputation [190, 200, 201], and structural equation models [202, 203, 204].

3.2.1 Regression calibration

Regression calibration is arguably the most simple and the most popular method to account for measurement error in regression models [118, 144, 161, 192, 134]. The basic idea behind this functional approach is to replace the observed exposure values with unbiased expected values

for true exposure [144, 205, 156]. The expected values for true exposure, $X_i(t)$, are typically modelled as a function of observed exposure $Z_i(t)$ and of other covariates $V_1, \dots V_p$. For instance, assuming a linear model in this context will lead to a measurement model of the form:

$$E(X_i(t)|Z_i(t), V_{i1}, \dots, V_{ip}) = \alpha_0 + \alpha_1 Z_i(t) + \sum_{k=1}^{p} \gamma_k V_{ik}, \tag{3.14}$$

where the unknown parameters α_0, α_1 and $\gamma_1, \cdots \gamma_p$ are typically estimated on a validation sample, i.e., a sample for which both true and observed exposure are measured accurately [127, 205]. Regression calibration leads to consistent estimates of the association between true exposure X_i and the outcome Y_i in linear regression. It is only approximately consistent in non-linear models, which are commonly used in epidemiological studies, such as logistic regression or proportional hazards models [206, 131, 207, 144, 156]. In proportional hazards models, regression calibration can result in moderate biases, even when the sample size is large [144]. Finally, regression calibration uses disjoint steps. It first estimates the conditional expected values of true exposure values $E(X_i(t)|Z_i(t))$. In a second step, these values are used as plug-in estimates for true exposure $X_i(t)$ when conducting statistical inference on the unknown risk parameters, implying that the uncertainty associated with the estimation of $E(X_i(t)|Z_i(t))$ is not accounted for in the estimation of these risk parameters. To remedy this problem, several authors have proposed to use bootstrap techniques in order to obtain standard errors and confidence intervals [208, 131, 209, 134], but it is questionable how common this option is in practice.

3.2.2 Simulation Extrapolation (SIMEX)

The aim of the simulation extrapolation approach (SIMEX), proposed by Cook et al. (1994) [210], is to correct for measurement error by establishing a relation between the magnitude of measurement error and the bias in parameter estimates. This relation is established via a series of simulations, in which different magnitudes of measurement error are added to the observed exposure data. For each simulation, the naive estimates for the parameter values are calculated. In a second step, a parametric function, called extrapolation function, is fitted to describe the relation between the magnitude of measurement error and the bias in these naive parameter estimates. This function is then used to extrapolate to the case of no measurement error. While SIMEX can be considered as robust in the sense that it does not rely on distributional assumptions on true exposure $X_i(t)$, its performance may critically depend on the correct specification of the extrapolation function [211]. As SIMEX consists of repeatedly fitting standard models, it is easy to implement, but it can be computer intensive [208]. Similarly to regression calibration, standard error estimates for this functional approach can be obtained via the bootstrap [210, 208]. However, due to the computational burden of the combination of SIMEX and the bootstrap to obtain standard error estimates, it is again questionable to what extent this option is chosen in practice [211].

3.2.3 Likelihood-based approaches

While regression calibration and SIMEX only result in consistent estimators in special cases, there are more elaborate methods, which are based on the likelihood that lead to fully consistent estimators [8]. These structural methods are more flexible than the functional methods

presented so far. Indeed, in order to account for classical measurement error using a likelihood-based approach, one merely has to specify a disease model, a measurement model and an exposure model as described in the beginning of this section. These models are then linked via conditional independence assumptions, in particular based on the assumption of non-differential measurement error, as defined in section 3.1.1. Likelihood-based approaches have been shown to be more efficient than regression calibration and SIMEX [196, 197, 199]. Moreover, they provide likelihood-based confidence intervals, which are more reliable than confidence intervals obtained for SIMEX or regression calibration [160], in particular in the case of non-linear models [8, 212]. However, in line with the well known "no free lunch" phenomenon [213, 214], this gain in efficiency comes at the cost of additional modelling assumptions [8]. In the case of classical measurement error, the full likelihood approach relies on the correct specification of the exposure model, i.e. of the distribution of true exposure $X_i(t)$ [215, 148], whereas functional methods for measurement correction at most make minimal distributional assumptions about $X_i(t)$. As true exposure $X_i(t)$ is by definition unobserved, it may be difficult to specify its distribution correctly, in particular in cases where no validation sample is obtainable. If the exposure model is misspecified, the full likelihood approach may be biased, contrary to SIMEX and regression calibration, which do not rely on an exposure model. Another major drawback of these methods is their computational burden [216, 199, 217, 162, 218]. In order to obtain the likelihood function to be maximised, one has to integrate the product of the measurement model, the disease model and the exposure model over the true and unobserved exposure values $X_i(t)$, which may be high-dimensional [217, 219]. Maximising the integrated likelihood may be challenging, mainly because it is rarely available in a closed form [217]. In a frequentist context, it is possible to use numerical methods, for instance the EM-Algorithm [220, 212, 132], or Gaussian quadrature [160, 8, 36] to maximise the integrated likelihood. However, the implementation of likelihood-based methods in complex situations may involve severe computational difficulties [149].

3.2.4 Methods based on multiple realisations of dose estimates

A recent approach to account for shared and unshared uncertainties in radiation epidemiology is to base statistical inference on risk estimates on multiple realisations of dose estimates for each cohort member [119, 107, 139]. Simon et al. (2015) [119] describe a Monte Carlo simulation design to generate possible dose realisations which explicitly separates unshared sources of uncertainty and sources of uncertainty that are shared between several cohort members. Several approaches have been proposed to integrate these realisations of dose estimates in risk estimation. In line with the idea of regression calibration, Little et al. (2014) [106, 221] have used an approach where multiple dose realisations are averaged and these averaged dose vectors are then used as plug-in estimates for true exposure in the disease model. A more elaborate approach was proposed by Stayner et al. (2007) [222]. The authors propose to calculate the likelihood for each possible dose realisation and a large number of pre-specified parameter values. These values are then averaged over all possible dose realisations to approximate the integrated likelihood and the maximum likelihood estimate can be derived by taking the maximum value of this integrated likelihood.

Chapter 4

The Bayesian hierarchical framework to account for uncertainty

While statisticians who chose the Bayesian approach to statistical inference still had to either hide their views from the statistical community or to be ready to staunchly defend their point of view some 50 years ago [223, 224, 225], there has been a remarkable surge in the popularity of this approach in the last decades. Indeed, as can be seen in Figure 4.1, when searching for the term "Bayesian" in PubMed, we can observe an exponential growth curve with the number of publications doubling about every five years since 1963. One of the main reasons for this

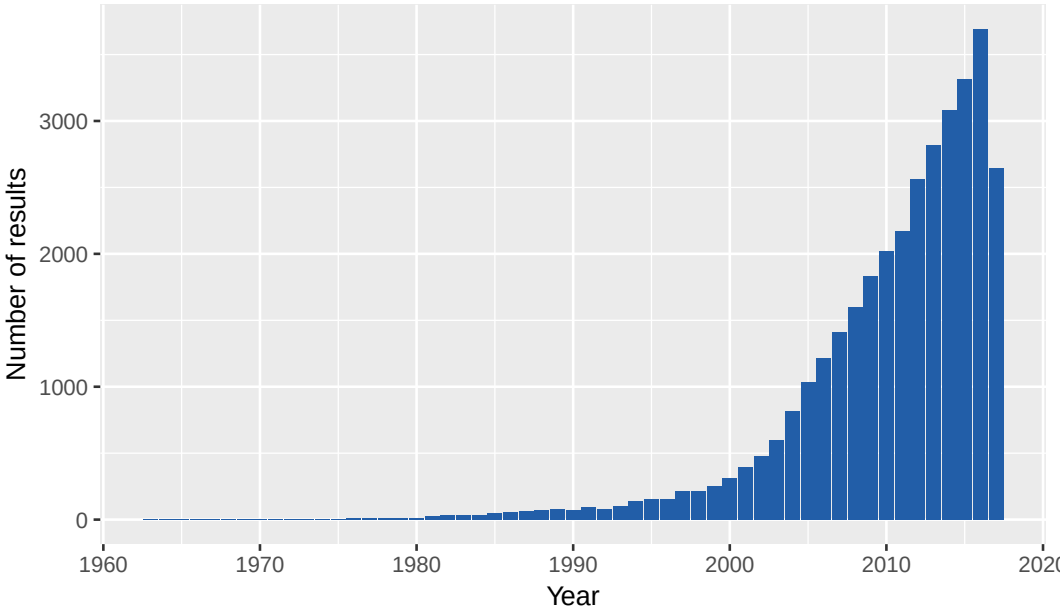

Figure 4.1: Number of results on PubMed when searching for the term "Bayesian" (results on the 5th of August 2017)

development is arguably the introduction of Markov Chain Monte Carlo (MCMC) methods to conduct Bayesian inference [226, 227] coupled with the availability of MCMC implementations in software packages (for instance WinBUGS in 1989) and technical improvements that led to a substantial increase in computing power [228, 229]. These improvements now make it possible to exploit the flexibility of the Bayesian approach by conducting inference for ever more

complex and realistic probabilistic models [230, 231, 232, 233, 234]. This chapter will give a brief overview of the Bayesian approach to statistical inference with a particular focus on its ability to account for exposure uncertainty in epidemiological studies and on the possibility to elicit prior distributions on unknown parameters based on expert knowledge.

4.1 History

There's no haters like Bayes' haters
They spit when they see a prior
Be careful when you offer your posterior
They'll try to kick it right through the door
But turn the other cheek it is not too sore
Of error they may yet tire!

— G. E. P. Box (who was married to R. A. Fisher's daughter Joan Fisher)

R. A. Fisher is often considered to be the founder of statistical inference [235, 236, 237, 238]. In line with this view, the Bayesian approach to statistical inference is often conceived as a relatively recent alternative to the frequentist approach (which is also referred to as classical statistics to stress this point). C. R. Rao goes so far as to say that before the introduction of the χ square test by Karl Pearson in 1900 and the t-test by Gosset in 1908, statistics meant observed data and descriptive summary figures [235]. However, this view does not really reflect the true sequence of events [239, 240, 241, 232, 242, 243, 244, 245, 246, 247, 248, 229]. Bayesian methods existed long before the frequentist approach to statistical inference, even though the term "Bayesian" is only in use since the 1950s. Before that time, what we call "Bayesian" today was known as inverse probability [243, 224].

Bayes and Laplace

The history of what is now called Bayesian inference starts with two men. The first was a clergyman who did not publish the one scientific paper for which he is famous today. The second, on the contrary, tried to publish as many scientific papers as possible in order to escape his father's will of becoming a clergyman [224]. The first is, of course, the Reverend Thomas Bayes. The second is Pierre-Simon Laplace. Many researchers today agree that Laplace's contributions to Bayesian statistics by far exceed the contributions of Bayes [249, 250, 241, 251, 246, 224]. Bayes' famous paper entitled "An Essay Towards Solving a Problem in the Doctrine of Chances" [252] was edited and published posthumously in 1763 by another Prebysterian minister called Richard Price. This paper only treats a special case of what is known as Bayes' theorem today, applied to a thought experiment involving a flat table and balls [249, 224]. Laplace, on the other hand, who was not aware of Bayes' publication [250, 246, 224], came up with the general form of Bayes' theorem and with its scientific application [224, 229]. Among other applications, Laplace used Bayes' theorem in astrophysics [240], to estimate sex ratios at birth for Paris and London in 1778 [253] and to estimate the size of the French population in 1802 [224]. It can therefore be argued that Bayesian statistics follows what is known as Stigler's law of eponymy

[254]. Stigler's law, which was first described by Merton (1957) [255] states that no scientific discovery is named after its original discoverer[1].

Fisher, Neyman and Pearson

What is nowadays referred to as the "frequentist" approach is in fact not a unique approach to statistical inference [240, 260, 261, 262, 263] but rather an amalgam of the ideas of R. A. Fisher on the one hand and Jerzy Neyman and Egon Pearson on the other hand [264, 265, 266] with which both sides would probably have been displeased [267, 243]. As a statistician and geneticist working on crop variance in Hertfordshire, England [268], Fisher promoted the ideas of the null hypothesis and the p-value in the 1920s in the framework of what he called "significance testing". Neyman and Pearson saw the need to base Fisher's ideas on a more mathematical foundation and set out to complete Fisher's approach in the 1930s, an endeavour which was met with very little cordiality by Fisher [269, 241, 261, 270, 271]. While agreeing on the fact that inverse probability, i.e., what is now called Bayesian statistics, was not a suitable approach to conduct statistical inference [272, 224], Fisher and Neyman agreed on few other points [241, 267]. Neyman and Pearson's theory of "hypothesis testing" featured not only a null hypothesis but also an alternative hypothesis, the concepts of error of the first and the second kind as well as statistical power [261]. According to Fisher, Neyman and Pearson's theory was founded on too strong prerequisites, thereby restricting the possibilities of application [270]. In his view, it might adequately be applied in the context of quality control and sampling inspection, but it would yield misleading results in scientific research [273, 274]. All in all, Fisher criticised Neyman and Pearson for being mere mathematicians lacking any contact with real scientific problems [241].

The renaissance of Bayesian statistics

After Neyman and Fisher had administered a "nearly lethal blow" [275] to Bayesian statistics, it took almost a quarter of a century for Bayesian ideas to emerge again [276]. By the late 1940s, frequentist teaching dominated most universities [224, 276]. While the use of Bayesian statistics in military applications had made spectacular successes during the Second World War, including Alan Turing's application in breaking the Enigma code, the documents describing these successes were mostly classified [224]. It is therefore not surprising that Dennis Lindley and Leonard J. Savage, who can be seen as two of the heroes associated with the "glorious Bayesian revival" of the 1950s and 1960s, did not set out with the intention of making Bayesian statistics respectable again. Both men merely tried to put traditional statistical techniques on a rigorous mathematical footing [224]. While they were able to build the logical foundations and to demonstrate the philosophical advantages of the Bayesian approach to statistical inference, in particular by showing its internal coherence [262, 231], the main problem of this Bayesian revival was that Bayesian methods seemed too complex for real problems. Savage, for instance, treated abstract questions on the most probable speed of neon light through beer and on the probability

[1]Examples of Stigler's law include Hubble's law of the continuing expansion of the universe, which was first discovered by Lemaître [256], the Poisson [257] and the Gaussian [258, 259] distribution, which were both introduced by De Moivre, the Metropolis algorithm, which will be described in section 4.2.3, the Pythagorean theorem [254] and, of course, Stigler's law itself.

that aspirin curls rabbits' ears [224], instead of tackling real world problems. According to Berger (2002) [277], for many years the general refrain was that "Bayesian analysis is nice conceptually; too bad it is not possible to compute Bayesian answers in realistic situations". Since the introduction of MCMC methods in the Bayesian toolbox, this situation has changed fundamentally [277]. Simpson et al. (2015) go so far as to state that today, applied researchers are only limited "by their data, their patience and their imaginations" when fitting increasingly complex models using the Bayesian approach to statistical inference [278].

The two sides of probability

The interpretation of the mathematical concept of probability plays an important role in the differences between the Bayesian and the frequentist approach to statistical inference [279]. The birth of this concept, which can be seen as a measure of uncertainty [224, 280, 281], is commonly traced to the year 1654 and more precisely to a correspondence between Blaise Pascal and Pierre de Fermat [282, 280, 283, 264]. Hacking (2006) [284] states that probability was congenitally Janus-faced. One side was epistemic and subjectivistic as a reasonable degree of belief. The other side was aleatory and objectivistic, associated with the long-run frequency of an event and the stochastic laws of chance processes [264, 284]. This conflated view of an objectivistic and a subjectivistic interpretation continued for James Bernoulli, Abraham De Moivre and Pierre Simon Laplace [280], who further elaborated the concept of mathematical probability. To them, it was natural to interpret the probability of an event at the same time as the "degree to which we should believe it will happen and the long-run frequency with which it does happen" [280]. According to Gigerenzer (1994) [264], it was Siméon-Denis Poisson, who first distinguished between the epistemic and the aleatory face of probability in print in 1837, thereby disrupting the unity of its two sides. The empiricist philosophers John Stuart Mill, Richard Leslie Ellis and Jakob Friedrick Fries subsequently decided to ban the epistemic face of probability [280], thereby obeying the general rule in positivism to reject all ideas and concepts that can not be observed empirically. Finally, in the 1920s and 1930s, Bruno de Finetti and Frank Ramsey introduced a behaviourist interpretation of the subjectivistic face of probability, thereby making it respectable within positivist philosophy [280]. In their view, the degree of belief of a person could be observed by their willingness to bet on these beliefs. Lindley, Savage and other proponents of the subjectivistic Bayesian movement base their interpretation of probability on de Finetti and Ramsey, while frequentism only accepts the purely objectivistic interpretation of probability.

In line with the two possible interpretations of probability, many authors choose to distinguish between epistemic and aleatory uncertainty [285]. In this view, epistemic uncertainty is due to a lack of knowledge and can be reduced by gathering more information. Aleatory uncertainty, on the other hand, is due to randomness and irreducible, no matter how much information we gather. Textbook examples for aleatory uncertainty include the tossing of dice, the drawing of cards from a shuffled pack and the sampling of a study participant from a population of interest. Lindley (2006) [251] proposes examples of epistemic uncertainty, including uncertainty on the two statements "The capital of Liberia is Morovia" and "The princes in the tower were murdered on the orders of Richard III". While both examples are undeniably due to a lack of knowledge, the uncertainty on the first statement can be easily reduced by searching for "Liberia" on the

internet, whereas it is not entirely possible to determine the veracity of the second statement. Indeed, this statement concerns an uncertain historical event, to be more precise, the disappearance of two young princes around 1483, who lived in the Tower of London and who may have put the legitimacy of the rulership of their uncle, Richard III, in doubt. In a Bayesian framework, both aleatory and epistemic uncertainty can be described by a common measure, namely by the mathematical concept of probability. In a frequentist context, on the other hand, only aleatory uncertainty can be described via this measure as this approach to statistical inference only accepts a purely objectivistic interpretation of probability.

It is futile to argue over the "correct" definition of probability, as a consistent mathematical theory of probability, as set out by Andrei Kolmogorov [280], does not depend on its interpretation [240, 281]. However, there are three important implications when one choses a purely objectivistic conception of probability. First of all, if probability is exclusively defined as the long-run frequency of an event under identical conditions, this ban of the subjectivistic side of probability simply reduces the types of problems that can be addressed via a probability model [240]. For instance, when an economist preparing the research budget for the US Air Force asked David Blackwell in 1950 to estimate the probability that a major war would occur in the next five years, Blackwell, who had not yet become a Bayesian, is reported to have replied, "Oh, that question just doesn't make sense. Probability applies to a long sequence of repeatable events, and this is clearly a unique situation. The probability is either 0 or 1, but we won't know for five years" [224]. Additionally, one has to be aware of the fact that the three seemingly innocuous words "under identical conditions" may actually conceal the epistemic side of probability that has been banned. For instance, when flipping a coin, it is generally acknowledged that there is a fifty percent chance for "heads" and a fifty percent chance for "tails". However, as pointed out by Loredo (1990) [240] and Diaconis et al. (2003) [286], the motion of a coin is adequately described by Newtonian mechanics. If we had exact information on the physical properties of the coin and on the initial conditions of the flip, we could predict the outcome with certainty. If we repeated the flip under the exact initial conditions, the outcome would always be the same[2]. Where does the randomness come from? Loredo (1990) argues that the randomness merely stems from our lack of knowledge of the initial conditions [240], thereby making the variability in the outcome of repeatable events inherently subjectivistic and epistemic. In a deterministic worldview where Laplace's demon could predict all the future of the universe without uncertainty if he had complete knowledge of the state of the universe at one time [287, 288], we can conclude that "there is no such thing as probability" [289], like de Finetti did when he proposed a purely subjectivistic interpretation of probability[290][3]. Finally, when banning the subjectivistic face of probability, one should be aware that this ban may give rise to misunderstandings and misinterpretations. In our everyday use of the word "probability", we generally accept that

[2]Diaconis even claims that his colleagues at the Harvard Physics Department built a coin flipper which always produces the same outcome [286].

[3]The deterministic view of Laplace brings us back to radioactive decay and nuclear physics, which we introduced in section 2.1. It can be argued that classical mechanics and thereby Laplace's demon was disproven by quantum mechanics and Heisenberg's uncertainty principle [288]. Anyways, as mentioned earlier, we can assume that the motion of a coin is adequately described by classical mechanics. Moreover, if the only objective probability concerns subatomic particles, we can at least say that if objective probability exists, it is negligible in our every day conception of the world.

humans describe something as being rather probable or improbable when referring to a degree of belief. For instance, during the Cold War, the probability of nuclear war was a widely debated subject. Similarly, today, we may want to discuss the probability that humans will be extinct in 200 years. In general, these probability statements are not understood to be grounded on the relative frequency of nuclear wars or human extinctions in parallel universes or when repeating the history of mankind a large number of times. They are understood to be an expression of our degree of belief.

4.2 An overview of the Bayesian approach to statistical inference

The main purpose of statistical inference is to derive information on unknown quantities of interest from observed data. Figure 4.2 illustrates the three basic ingredients that are common to parametric statistical inference, namely the data, a vector of unknown parameters $\boldsymbol{\theta}$, and a probability model (also referred to as "likelihood"), which creates a link between the two of them. As the aim of most epidemiological studies is to assess the association between a disease outcome and a certain number of risk factors, we will focus in this section on the situation, where the occurrence of a disease outcome $Y = (Y_1, Y_2, \ldots, Y_n)$ observed for n individuals follows a probability distribution depending on a vector of unknown parameters $\boldsymbol{\theta}$ and an exposure variable $X = (X_1, X_2, \ldots, X_n)$[4], which is commonly considered as known and non-random. Moreover,

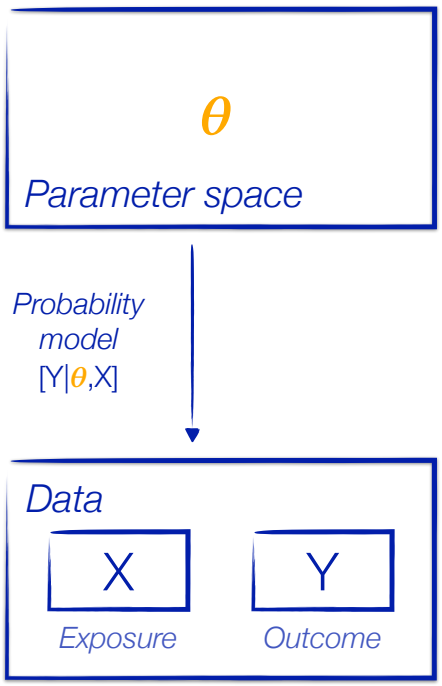

Figure 4.2: The basic components of parametric statistical inference

we will adopt the notation introduced by Gelfand (1990) [291], where probability distributions

[4]For the sake of simplicity, we will focus on one exposure variable in the following, but the reasoning can easily be extended to several exposure variables.

are denoted by square brackets[5]. The probability model, or likelihood, denoted as $[Y|\boldsymbol{\theta}, X]$, describes the conditional probability distribution of the outcome Y given the exposure X and the vector of unknown parameters $\boldsymbol{\theta}$. On the one hand, if we knew the value of the unknown parameters $\boldsymbol{\theta}$, we could use this model to generate possible realisations of the outcome Y. The variability of these realisations is thereby described by the assumed probability model. This simulation process can be seen as a forward propagation of uncertainty from $\boldsymbol{\theta}$ to Y. The aim of statistical inference, on the other hand, is a backward propagation of uncertainty: Given the outcome Y and the likelihood $[Y|\boldsymbol{\theta}, X]$, we want to estimate the value of the unknown parameter $\boldsymbol{\theta}$ and quantify the uncertainty associated with this estimation. In this sense, statistical inference can be conceived as an inversion process [232, 292].

4.2.1 Using Bayes' theorem to solve the inversion problem of statistical inference

From the Bayesian point of view, the inversion problem of statistical inference can be resolved by simply reversing the sense of the probabilistic conditioning, as illustrated in Figure 4.3. In this framework, the uncertainty on the estimated parameters is described by the conditional probability distribution of these parameters $\boldsymbol{\theta}$ given the observed data, which is also called posterior distribution $[\boldsymbol{\theta}|Y, X]$. The posterior probability distribution (often simply referred

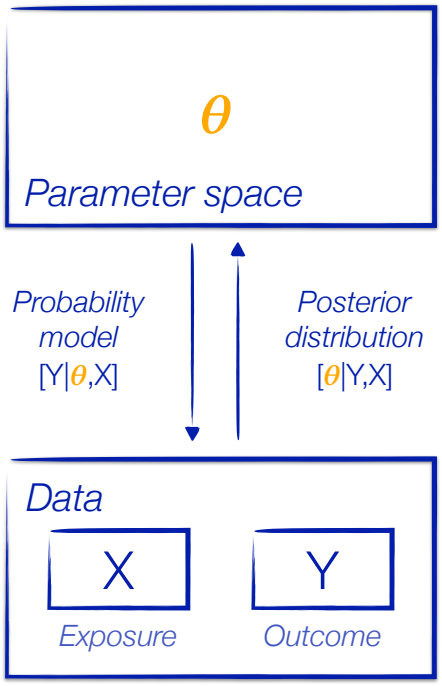

Figure 4.3: The Bayesian solution to the inversion problem of statistical inference

to as posterior or posterior distribution) can be obtained by applying Bayes' theorem in the

[5]In this notation $[X, Y]$ denotes the joint probability distribution of the two random variables X and Y, $[Y|X]$ is the conditional probability distribution of Y given X and $[Y]$ is the marginal probability distribution of Y. $[X]$ denotes a probability distribution for discrete random variables and a probability density for continuous random variables.

following way:

$$[\boldsymbol{\theta}|Y,X] = \frac{[\boldsymbol{\theta},Y|X]}{[Y|X]} = \frac{[\boldsymbol{\theta}]\cdot[Y|\boldsymbol{\theta},X]}{[Y|X]} \tag{4.1}$$

where $[\theta]$ is the prior probability distribution (often simply referred to as prior). The prior is assigned to the vector of parameters $\boldsymbol{\theta}$ to describe our state of knowledge (or in other words our epistemic uncertainty) about $\boldsymbol{\theta}$ before analysing the data. It is important to note that the fact that the values of $\boldsymbol{\theta}$ are described by a probability distribution does not imply that the true values of $\boldsymbol{\theta}$ are considered to be random. Bayesians generally assume that the values of $\boldsymbol{\theta}$ are fixed and unknown and that it is merely the epistemic uncertainty that we have on these values that can be described by a probability distribution. The last expression implicitly assumes that $[\boldsymbol{\theta}|X] = [\boldsymbol{\theta}]$, i.e. our prior uncertainty on the unknown parameters does not depend on the exposure values X. $[Y|X]$ can in theory be obtained by integrating the joint distribution $[\boldsymbol{\theta},Y|X]$ over all possible values of $\boldsymbol{\theta}$. In practice, it is often treated as a superfluous normalising constant, as it does not depend on the values of the unknown parameters $\boldsymbol{\theta}$ [281]. In this case, the recipe to conduct Bayesian inference becomes:

$$\begin{aligned}[\boldsymbol{\theta}|Y,X] &\propto [\boldsymbol{\theta},Y|X] \\ &\propto [\boldsymbol{\theta}]\cdot[Y|\boldsymbol{\theta},X].\end{aligned} \tag{4.2}$$

The quantiles of the resulting posterior distribution $[\boldsymbol{\theta}|Y,X]$ can then be used to express the uncertainty associated with the estimation of the unknown vector of parameters $\boldsymbol{\theta}$. Thereby, a Bayesian 95% credible interval (corresponding to the 2.5th and the 97.5th percentile of the posterior distribution) can be interpreted in the following way: Given the data we observed, there is a 95% chance that the true value of the unknown parameter $\boldsymbol{\theta}$ lies in the obtained credible interval.

It is common to present Bayesian statistics as an inductive approach to statistical inference in which we learn about the general from the particulars [293]. In this view, Bayesian inference is described as a smooth learning process, where the prior uncertainty about the vector of unknown parameters, $[\boldsymbol{\theta}]$, is updated through the information provided by the data (and described by the likelihood $[Y|\boldsymbol{\theta},X]$) to obtain the posterior distribution $[\boldsymbol{\theta}|Y,X]$, which reflects the uncertainty about $\boldsymbol{\theta}$ after observing the data. In contrast, the frequentist approach, which we will present in more detail in the next section, is associated with hypothetico-deductive inference, as advocated by Karl Popper[6]. Unfortunately, the description of Bayesian inference as an inductive learning process is somewhat reductionist, as one tends to forget that Bayesians dispose of excellent tools for the testing of hypotheses and for model checking [294, 293]. More importantly, it can even lead to the false impression that "Bayesianism means never having to say you're wrong" [295]. In order to avoid such an impression, we preferred to emphasise in this section that Bayesian inference provides us with a unique solution to the inversion problem of statistical inference and that the epistemic uncertainty about the fixed and unknown vector of parameters $\boldsymbol{\theta}$ of interest can be expressed via a probability distribution. The choice of this prior probability distribution, which is a crucial point when conducting Bayesian inference, will be discussed in section 4.3.

[6] According to this hypothetico-deductive approach, researchers start with general theories to deduce hypotheses which can be tested on well-designed experiments. Contrary to inductive inference, deductive inference moves from the general to the particulars. One of the main ideas of deductive inference is that hypotheses can never be accepted, but only rejected or falsified [293].

4.2.2 Comparison with the frequentist approach to statistical inference

The frequentist approach to statistical inference is based on a one-sided interpretation of probability, where probability is solely understood as the long-run frequency of an event that can be repeated an infinite number of times under identical conditions [143]. In this framework, it therefore makes no sense to assign a probability distribution to an unknown parameter and to resolve the inversion problem of statistical inference by using Bayes' theorem. To quantify their

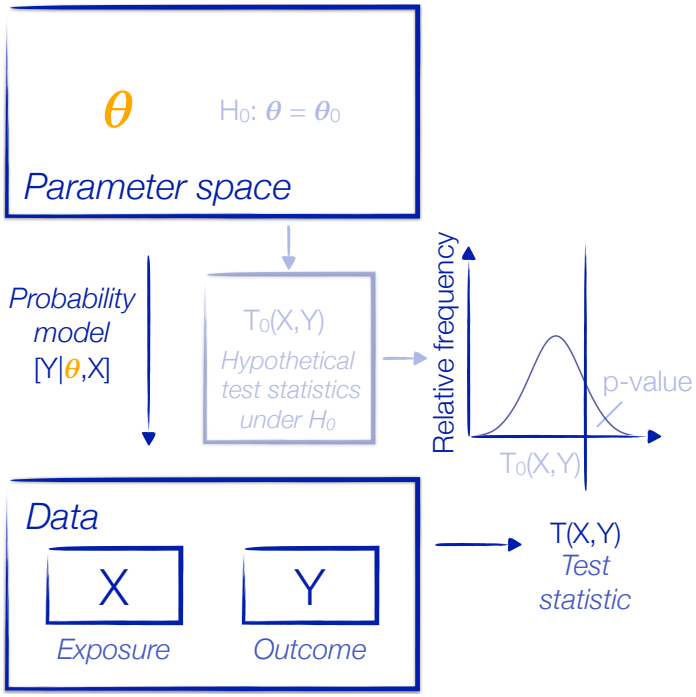

Figure 4.4: A sketch of the frequentist approach to statistical inference

uncertainty in the estimation of the unknown parameters when conducting statistical inference, frequentists therefore generally rely on three additional concepts: A null hypothesis H_0, a test statistic $T(X, Y)$ and the possibility to repeat the data generation process a large number of times. The test statistic $T(X, Y)$ is some function of the observed data. The main idea of frequentist inference is to construct a hypothetical infinite population under the assumption that the null hypothesis is true [270, 238] and to derive the distribution of the test statistic obtained for random samples drawn from this population. This distribution can be expressed by imagining a large number of realisations of the test statistic under the null hypothesis, which we can denote as $T_0(X, Y)$. As illustrated in Figure 4.4, the p-value can then be seen as the relative frequency with which these hypothetical test statistics, which could be produced under the null hypothesis, would exceed the value of the observed test statistic $T(X, Y)$. Put differently, it tells us how many times we would observe a result which is more extreme than the one that we actually observed if we were in a world in which the null hypothesis is true and we repeated our data collection a large number of times. Similarly, frequentist confidence regions are based on the hypothetical repeatability of the data analysis. In accordance, the interpretation of a frequentist 95% confidence interval is the following: If we repeated our analysis on an infinite number of independent data sets that are subject to the same data generation mechanism as the present data set, 95% of these intervals would cover the true value of the unknown parameter $\boldsymbol{\theta}$.

Contrary to the Bayesian approach, which has a unique way of deriving the posterior distribution given the data, the prior and the probability model, there is no unique way to derive a test statistic $T(X, Y)$ for a given probability model and data set in frequentist inference [292]. However, there are a number of principles and desirable properties of $T(X, Y)$ to guide its choice, including unbiasedness, sufficiency, consistency and efficacy [296, 240, 262, 297].

In summary, the main difference between Bayesian and frequentist statistics is that Bayesians can make probability statements about the data given the parameters and about the parameters given the data, while frequentists choose to confine themselves to probability statements about the data given the parameters. Moreover, the frequentist approach relies on a null hypothesis (and an alternative hypothesis in the case of the theory of "hypothesis testings" by Neyman and Pearson) and the idea that the data generation could be repeated a large number of times. Bayesian inference, on the other hand, relies on the specification of a prior distribution in order to derive the posterior distribution.

4.2.3 Markov chain Monte Carlo methods

Markov Chain Monte Carlo (MCMC) methods are a class of algorithms that provide a means of drawing samples from complex probability distributions which only have to be known up to a normalising constant [232, 298]. They achieve this sampling by constructing a discrete-time Markov chain[7], which is designed so as to generate random draws from a potentially high-dimensional target distribution [299, 300]. When MCMC methods are used to conduct Bayesian inference, this target distribution is typically the posterior distribution. Based on these samples, it is straightforward to estimate various quantities, including the posterior mean and the posterior median, but also the probability that the true parameters are greater than a pre-specified threshold, for instance.

The origins of MCMC algorithms and more generally of modern Monte Carlo methods can be found in the statistical physics literature [301, 302, 303, 227], and in particular in research projects involving the development of the atomic bomb and the first nuclear reactors [233, 304]. Being therefore inextricably linked to the history of uranium mining, some may think that it would have been more adequate to present MCMC methods in chapter 3. However, and in spite of the fact that after their introduction it took almost 40 years for MCMC methods to penetrate statistical practice [248, 224, 227], these methods have had a profound and far reaching influence on both Bayesian methodology and its application [299, 233]. According to McGrayne (2011) the introduction of MCMC even led Adrian F. M. Smith, who had been a student of Lindley, to quit statistics and to become administrator of the University of London, only three years after becoming the first Bayesian president of the Royal Statistical Society in 1995. He justified his career choice by simply saying that with the Bayesian paradigm and MCMC all problems of statistics had been solved [224].

As the necessary conditions for the convergence to a target probability distribution of interest are given by Markov chain theory (see for instance Roberts et al (1994) [305] or Robert et

[7] A Markov chain is a sequence of random variables $X_0, X_1, \ldots, X_t, X_{t+1}, \ldots$ that satisfies the so-called Markov property according to which the conditional probability distribution of future states X^{t+1} depends only on the past via its present state X^t: $[X^{t+1}|X^t, X^{t-1}, X^{t-2}, ..., X^1, X^0] = [X^{t+1}|X^t]$.

al. (2002) [231]), MCMC methods provide a generic approach to conduct Bayesian inference. Thereby, the use of these methods allow to focus on the development of models that realistically reflect the problem instead of having to concentrate on proofs of the desirable properties of a proposed estimator as can be the case for frequentist inference (see section 4.2.2). Moreover, for MCMC methods, the asymptotic convergence of the generated values to the target distribution concerns the number of iterations of the Markov chain that tend to infinity whereas the asymptotic properties of frequentist estimates concern the number of study participants in a sample. While both an infinite number of MCMC draws and an infinite sample of study participants are impossible to achieve in practice, in most studies, it appears to be easier to increase the number of MCMC iterations than to increase the number of study participants, in particular with technical improvements leading to increases in computer power. Finally, a remarkable aspect of MCMC methods is that, once the Markov chain has converged to the target probability distribution, each iteration of the algorithm provides us with a realisation of this target distribution. In this sense they can be seen as more efficient than other Monte Carlo methods such as rejection sampling where several iterations could be necessary to obtain one realisation of the target distribution. However, contrary to the latter method, the draws produced by an MCMC algorithm are not independent and may present a strong autocorrelation. Moreover, particular attention should be paid to the convergence of a Markov chain to its target distribution. There are no diagnostic tools that can assure the convergence of the chain, but merely tools to detect divergence. A common way to detect a potential divergence is to generate multiple Markov chains with dispersed starting values and to compare the between and the within variability of the values produced by the different chains as well as the extent to which they visit regions of low posterior probability [299]. Gelman et al. [306] suggest a diagnostic to assess this difference between the estimated between-chains and within-chain variances. This diagnostic is based on the ratio between the pooled variability by gathering all chains and the within-chain variability. A ratio close to 1 indicates that a sufficient burnin was removed (omitted first iterations to forget starting values) and that all chains visit the whole of the sample space. The most popular MCMC methods are the Metropolis-Hastings algorithm and the Gibbs sampler, which will be briefly presented in the following.

The Metropolis-Hastings algorithm

The origins of MCMC methods can be found in the Metropolis algorithm, introduced in the paper "Equation of State Calculations by Fast Computing Machines" [307], which is arguably the most influential publication in the history of computational physics [308], published in the Journal of Chemical Physics in 1953. According to Shonkwiler et al (2009), the Metropolis algorithm has even been cited among the top ten algorithms having the greatest influence on the development of science and engineering [304]. As mentioned in section 4.1, the Metropolis algorithm can be seen as yet another example of Stigler's law of eponymy. Indeed, the only real contribution Metropolis made to the development of the algorithm for which he is famous today is computer time [309, 308][8]. In the beginning of the 1950s, he was in charge of a large computing machine in Los Alamos, called the MANIAC (Mathematical Analyzer, Integrator And Computer). On this computer, researchers like Enrico Fermi, Stan Ulam and John von Neumann

[8]Contrary to the Monte Carlo method, for which he at least suggested the name [302].

made important scientific developments [309]. In particular, Metropolis collaborated with the two married couples Marshall and Arianna Rosenbluth and Edward and Augusta Teller, who developed what is today known as the Metropolis algorithm. As the contributor of computer power, Metropolis co-authored the publication describing this algorithm, but he neither came up with its idea nor did he participate in its implementation [308]. His fame today is therefore mainly due to a convention on the alphabetical ordering of co-authors in the physical literature [310, 311] and if Augusta Teller had kept her maiden name, the Metropolis algorithm would probably be known as Harkányi algorithm today.
Almost twenty years after its introduction in the physical literature, the Metropolis algorithm was generalised by Hastings (1970) in a paper in Biometrika [312], but the potential of MCMC methods for Bayesian inference was not realised until Gelfand and Smith (1990) [291] demonstrated its usefulness in conducting Bayesian inference [232, 227].

When conducting Bayesian inference via the Metropolis-Hastings algorithm, given a current value $\boldsymbol{\theta}^{\text{t-1}}$, at each iteration t, the next state of the chain is chosen by first generating a candidate value $\boldsymbol{\theta}^{\text{cand}}$ from a proposal distribution $[\boldsymbol{\theta}^{\text{cand}}|\boldsymbol{\theta}^{\text{t-1}}]$. This candidate state is then either accepted or rejected with a certain probability. The acceptance probability is given by

$$\rho(\boldsymbol{\theta}^{t-1}, \boldsymbol{\theta}^{\text{cand}}) = \min\left\{ \frac{[\boldsymbol{\theta}^{\text{cand}}|Y,X]}{[\boldsymbol{\theta}^{\text{t-1}}|Y,X]} \frac{[\boldsymbol{\theta}^{t-1}|\boldsymbol{\theta}^{\text{cand}}]}{[\boldsymbol{\theta}^{\text{cand}}|\boldsymbol{\theta}^{\text{t-1}}]}, 1 \right\} \tag{4.3}$$

where $[\boldsymbol{\theta}|Y,X]$ is the posterior distribution evaluated at $\boldsymbol{\theta}$. If $\boldsymbol{\theta}^{\text{cand}}$ is accepted $\boldsymbol{\theta}^{\text{t}} = \boldsymbol{\theta}^{\text{cand}}$. Otherwise $\boldsymbol{\theta}^{\text{t}} = \boldsymbol{\theta}^{\text{t-1}}$.
An iteration t of the Metropolis-Hasting algorithm can be written in the following way:

$$\boldsymbol{\theta}^{\text{cand}} \sim [\boldsymbol{\theta}|\boldsymbol{\theta}^{\text{t-1}}]$$

$$\boldsymbol{\theta}^{\text{t}} = \begin{cases} \boldsymbol{\theta}^{\text{cand}} & \text{with probability } \rho(\boldsymbol{\theta}^{\text{t-1}}, \boldsymbol{\theta}^{\text{cand}}) \\ \boldsymbol{\theta}^{\text{t-1}} & \text{otherwise} \end{cases}.$$

The successful design of a Metropolis-Hastings algorithm critically depends on the appropriate choice of the proposal distribution $[\boldsymbol{\theta}^{\text{cand}}|\boldsymbol{\theta}^{\text{t-1}}]$ according to which the candidate values $\boldsymbol{\theta}^{\text{cand}}$ are generated at each iteration [313]. Firstly, it is important that this probability distribution is known analytically, at least up to a normalising constant, and that it is possible to draw samples from this distribution in a time-efficient way. Moreover, the proposal distribution has to meet certain minimum requirements to make sure that the limiting distribution of the Markov chain produced by the algorithm coincides with the target distribution of interest, i.e. the posterior distribution. In the case of a symmetric proposal distribution, the ratio given in (4.3) simplifies and becomes the ratio of the posterior distribution evaluated at $\boldsymbol{\theta}^{\text{cand}}$ and $\boldsymbol{\theta}^{\text{t-1}}$.
It is interesting to note that the algorithm systematically accepts candidate values for which the ratio $\frac{[\theta^{\text{cand}}|Y,X]}{[\theta^{\text{cand}}|\theta^{\text{t-1}}]}$ is greater or equal to the corresponding ratio for the current value $\frac{[\theta^{\text{t-1}}|Y,X]}{[\theta^{\text{t-1}}|\theta^{\text{cand}}]}$. More importantly, a candidate value can also be accepted if this is not the case, i.e. if there is a decrease in this ratio for the candidate value compared to the current value of the chain. The acceptance probability (4.3) is inversely proportional to the magnitude of this decrease. This property allows the chain to visit the tails of the target distribution. It is very important to monitor the acceptance rate of the algorithm (i.e. the proportion of iterations for which the proposed candidate value is accepted). In particular, this quantity can be used to guide the

calibration of the proposal distribution, for instance, by increasing or by reducing its variability. There are different recommendations concerning the optimal acceptance rate with a general consensus that this quantity should be around 0.40. There are two alternative strategies to attain an optimal acceptance rate. One can either use the algorithm after an initial calibration phase or use an adaptive version of the algorithm. Instead of updating the whole vector of unknown parameters $\boldsymbol{\theta}$ in one step, it is often more convenient to divide $\boldsymbol{\theta}$ in components and to update these components one by one [299, 313].

Gibbs sampling

Gibbs sampling can be considered as a special case of the general Metropolis-Hastings algorithm [314, 299, 231], or, to be more precise, as a combination of Metropolis-Hastings algorithms on different components. In this special case, which is based on a component-wise updating of the unknown parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, ...\theta_p)$, the proposal distribution for each component θ_k is its full conditional distribution $[\theta_k^t|\theta_1^t, ...\theta_{k-1}^t, \theta_{k+1}^{t-1}, ...\theta_p^{t-1}]$. An iteration t of Gibbs sampling algorithm can be written in the following way:

$$[\theta_1^t] \sim [\theta_1|\theta_2^{t-1}, \ldots, \theta_p^{t-1}]$$

$$[\theta_2^t] \sim [\theta_2|\theta_1^t, \theta_3^{t-1} \ldots, \theta_p^{t-1}]$$

$$\ldots$$

$$[\theta_p^t] \sim [\theta_p|\theta_1^t, \ldots, \theta_{p-1}^t].$$

As the acceptance probability in this case is one, the Gibbs sampler reduces to the sequential sampling from the full conditional distributions. In the context of this algorithm, the discussion concerning the optimality of acceptance rates is therefore irrelevant. This algorithm presupposes that the full conditional distributions are calculable and that it is possible to draw samples from these distributions. In cases where it is not possible to draw samples from some of the full conditional distributions, it is possible to replace the updating of the corresponding component by a classical Metropolis-Hastings step. This version of the algorithm is referred to as Metropolis-within-Gibbs-sampling. In cases where the full conditional distributions are available, both the general Metropolis-Hastings algorithm and Gibbs sampling are possible. On the first glance, one might believe that Gibbs sampling is always preferable to a general Metropolis-Hastings implementation, because in the context of Gibbs sampling all candidate values are accepted and because it avoids the somewhat arbitrary choice of a proposal distribution. However, this intuition may be wrong. In particular, a poor parameterisation can greatly prolong the time until the convergence to the target distribution of the Gibbs sampler. This problem can arise for instance when conducting inference on mixture models where one or several components of the mixture consist of only few observations.

4.3 The choice of the prior distribution

There's no Theorem like Bayes theorem
Like no theorem we know
Everything about it is appealing
Everything about it is a wow
Let out all that a priori feeling
You've been concealing up to now.
—— George E. P. Box.

Since the introduction of Bayesian inference, there has been an ongoing debate about whether the need to specify a prior distribution in the Bayesian approach is a bug or a feature. On the one hand, the Bayesian approach is often praised for the ability to incorporate prior information in statistical inference in a natural way [195, 143, 315, 316], as the use of prior information may improve the accuracy and the reliability of the results by borrowing strength from previous analyses or from expert knowledge [317, 318, 319, 320, 321, 144]. On the other hand, the need to specify a prior distribution is probably the most criticised point in the Bayesian framework [292] and therefore the most important reason for not using this approach to statistical inference [322, 248, 281, 253, 234]. Indeed, opponents of the Bayesian approach to statistical inference are often uncomfortable with the seemingly arbitrariness of the choice of the prior distribution [323, 319]. Moreover, the possibility of a subjectivistic and epistemic interpretation of probability seems to contradict the ideal of the objectivity of scientific reasoning [239, 324, 319, 318, 281, 253]. Even Bayesians have described the necessity to define a prior distribution as "vexing" [325] and as a "sticking point" [326].

4.3.1 Prior choices that avoid the use of external information

In light of the difficulties involved in the specification of informative prior distributions, many applied researchers choose a Bayesian approach which explicitly excludes all available prior information [327, 328, 323, 277, 245, 315]. These prior distributions are in general constructed with the aim to base statistical inference only on the assumed probability model and the observed data [329] and to express complete ignorance on the values of the parameters a priori. The underlying idea is to use Bayesian techniques while "letting the data speak for themselves". Unfortunately, there is no general agreement on the mathematical expression of ignorance via a probability distribution [323, 326, 243, 330, 331]. An intuitive choice is to use a flat prior distribution, for instance a uniform distribution defined on a large interval or a normal distribution with a very large variance. Indeed, these choices imply that a large range of possible parameter values are equally probable a priori (in the case of the uniform distribution) or at least that a large range of parameter values have a similar probability a priori (in the case of the normal distribution). In line with this intuition, Bayes and Laplace both mostly relied on uniform prior distributions [277, 245, 224]. However, opponents of the Bayesian approach, including Fisher [272], have legitimately criticised the choice of vague prior distributions on the grounds that they are not robust to parameter transformations [328, 332, 333]. Based on this criticism, Harold

Jeffreys developed the concept of "noninformative" prior distributions. These noninformative prior distributions, which are based on the Fisher information, are robust to parameter transformation [334, 234]. Similarly, Bernardo (1979) [324] introduced the idea of a reference prior which is also robust to parameter transformation and based on information-theoretic considerations. However, both prior choices may turn out to be improper, i.e. they result in a prior distribution that does not integrate to a finite number [323, 335]. Moreover, despite their objective of having little or no impact on inference, non-informative prior distributions can be highly informative and even overwhelm the data [323, 336, 228]. They can also lead to improper posterior distributions [328]. These problems arise in particular for variance parameters [337, 338], when the sample size is small [228, 320, 321] and when the posterior distribution is high-dimensional [323]. Gelman (2013) argues that this phenomenon can be explained by the fact that vague and noninformative prior distributions put too much probability on extreme values, which are never going to be plausible, thereby disturbing the posterior probabilities [339].
Finally, another common strategy to avoid the use of external information in the derivation of the prior distribution is the use of empirical Bayes techniques, where the parameters of the prior distribution are estimated by using the observed data [292, 340].

4.3.2 Informative prior distributions

There are a number of authors who advocate the use of informative prior distributions [341, 317, 262, 319, 342, 343]. As mentioned above, the integration of external information in the form of prior distribution can increase the precision in the estimation of unknown parameters and improve predictive accuracy of the outcome Y [319, 315, 321]. Moreover, Albert et al. (2012) point out that when constructing complex probability models, there may be sub-models that are well informed by the observed data and other sub-models for which it is necessary to use expert opinion to supplement the information provided by the data [344]. In this vein, when using complex models with a large number of uncertain input quantities, for instance in climate modelling, the subjective Bayesian approach appears almost as the only feasible option [319], synthesising historical data, the observed data and expert opinion in a coherent framework.
Finally, Stephen Senn (2011) [345] argues in his paper "You may believe you are a Bayesian but you are probably wrong" that there are a number of appealing features that are often brought forward when using Bayesian techniques, but which are only valid when a subjective Bayesian approach is used. For instance, the Bayesian hierarchical approach to account for measurement error, which we will present in the next section, can be praised for being a coherent approach to account for both exposure and parameter uncertainty. However, as pointed out by Senn, "the degree of uncertainty must be determined by the degree of certainty and certainty has to be a matter of belief so that it is hard to see how prior distributions that do not incorporate what one believes can be adequate for the purpose of reflecting certainty and uncertainty" [345]. Senn also argues that subjective Bayesianism is perfect in theory, but difficult to apply in practice. In particular, to him "Bayesian theory is a theory of how to remain perfect but it does not explain how to become good" [345]. Despite the advantages of the subjective Bayesian approach, it is rarely used in scientific applications. The main reasons for the reluctance to use a subjective perspective are maybe the idea to ensure a certain "objectivity" of scientific analyses [239, 324, 318, 319, 343, 281, 253] and a lack of domain knowledge. An appealing solution to these

two problems of the subjective Bayesian approach is to specify informative prior distributions through expert prior elicitation, where probability distributions on the unknown parameters are derived based on expert knowledge. If the elicitation of expert knowledge is performed in a transparent and repeatable manner, it can be argued that it allows to derive prior distributions informed by experts in the domain in an objective manner. Section 4.3.3 will give a brief overview of the literature concerning the elicitation of prior distributions based on expert knowledge.

4.3.3 The elicitation of prior distributions by expert knowledge

While the translation of the judgements of experts into probability distributions plays an important role when the aim is to derive informative priors in the Bayesian framework, expert elicitation is far from being limited to this application. Indeed, methods for the elicitation of expert knowledge are also used to define uncertainty on input parameters in mechanistic models which are used to predict complex processes, for instance, climate change, the risk of an accident in a nuclear installation or to include parameter uncertainty in other risk analyses [346, 347, 279, 348, 349, 350, 351, 352, 353, 354]. Despite this broad range of applications, there is no consensus in the literature on the best way to elicit a prior distribution. It is generally acknowledged that it is difficult to translate uncertainty in a probability distribution and that there are a number of cognitive biases, which we will briefly discuss in the next section, which should be accounted for in the design of an elicitation task. Consequently, we will summarise some general recommendations for the design of elicitation tasks. Moreover, it is challenging to derive informative prior distributions based on the elicited data, in particular when it comes to the combination of the opinion of several experts. Various approaches exist in this context but there is no general consensus on the approach to adopt in given situation. We will describe and compare several approaches for the combination of expert opinion in chapter 5.

Cognitive biases in expert prior elicitation

Research on expert elicitation has been strongly influenced by a series of papers published by Tversky and Kahneman [355, 356] in the early 1970s in the context of their "heuristics and biases" research program in cognitive psychology [357, 264, 358, 359, 347, 360, 361]. In their research program, Tversky and Kahneman described human inadequacies when assessing probabilities and asserted that humans made probability judgements based on heuristics [360]. These heuristics, which are common-sense or rule of thumb decisions based on experience [359], do not necessarily comply with probability theory and may lead to systematic and predictable bias in probability judgements. They include availability, adjustment and anchoring, and overconfidence [279]. The availability heuristic describes the fact that assessors tend to be influenced in their probability judgments by the frequency with which they can recall an event [359, 279]. According to Wolfson et al. (2016), assessors will for instance overestimate the probability to die from causes receiving a lot of media attention, such as botulism, and to underestimate the probability to die from more common but less notorious causes such as stomach cancer [361]. According to the adjustment and anchoring heuristic, assessors tend to start with an initial estimate, or anchor value, and then to adjust this initial estimate when asked to estimate a quantity or to assess an uncertainty [279]. This initial estimate is then adjusted up or down. However,

these further adjustments are often insufficient, resulting in an assessment that is biased towards the anchor value [359, 361]. Finally, overconfidence describes the fact that assessors tend to be more confident about their individual answers than their overall number of correct answers suggests [360].

General recommendations for the design of an elicitation task

In light of the possible biases that may arise in the elicitation of prior distributions, there is a general consensus in the literature that the elicitation of prior information based on expert knowledge is essentially difficult [357, 347] and that no straightforward reliable procedure is yet available [361]. However, there is evidence that experts perform better at elicitation tasks than non-experts [279, 361] and there are a number of general recommendations which can minimise the impact of cognitive biases. The most fundamental principle is maybe that experts should only be asked questions that they can understand and that they can answer [343]. While this principle may seem self-evident and superfluous on the first glance, it entails a certain number of constraints. First of all, it implies that experts should only be asked to provide information on quantities that are observable [358, 350]. Moreover, they should not be asked to quantify the variance or other moments of a distribution [230], except possibly the first moment, i.e. the mean [358, 361]. More generally, it is recommended to ask for the quantiles of a distribution rather than asking for its moments [362, 279, 363, 361]. In principle, the quantiles of a distribution can be elicited in a direct or in an indirect manner. Elicitation methods that infer the probability distribution of an uncertain variable via comparative judgments are generally referred to as indirect methods whereas direct methods ask experts to explicitly state the moments or the quantiles of a probability distribution [279, 351]. According to Budescu et al. (2011), indirect methods are considered more natural and less demanding cognitively than direct elicitation methods [351]. One of the most popular choices when using an indirect method for the elicitation of expert knowledge is via a thought experiment in which experts can either spin a fortune wheel or make a bet on the values of a quantity of interest [279, 1, 351]. For instance, Abbas et al. (2008) [1] asked students in an online questionnaire to assess the high temperature in Pablo Alto on a fixed day in the following week. In this online questionnaire, the students were confronted with the screen shown in Figure 4.5, where they could either choose to spin the fortune wheel presented on the left or make a bet on the maximum temperature in Pablo Alto on the 12th of December 2006.

Finally, a strategy to avoid heuristics and biases in the design of an elicitation task may be to ask about the observed variability concerning several realisations of an uncertain variable rather than to ask questions about the probability of values that could be observed for a single realisation of this variable. In line with this, several experiments showed that the cognitive biases and the heuristics identified by Tversky and Kahneman can disappear when probability statements are formulated in a frequency format [364, 365, 366, 367]. Fischer et al. (2015) [76], for instance, employed methods for the elicitation of expert knowledge to develop a job-exposure matrix by confronting experts with the hypothetical experiment of assembling 100 workers in a room who are asked to raise their hand if they have experienced a certain exposure condition during their working career. Finally, there is a general consensus on the fact that the elicitation of the quantities of interest should be preceded by a training phase and that frequent

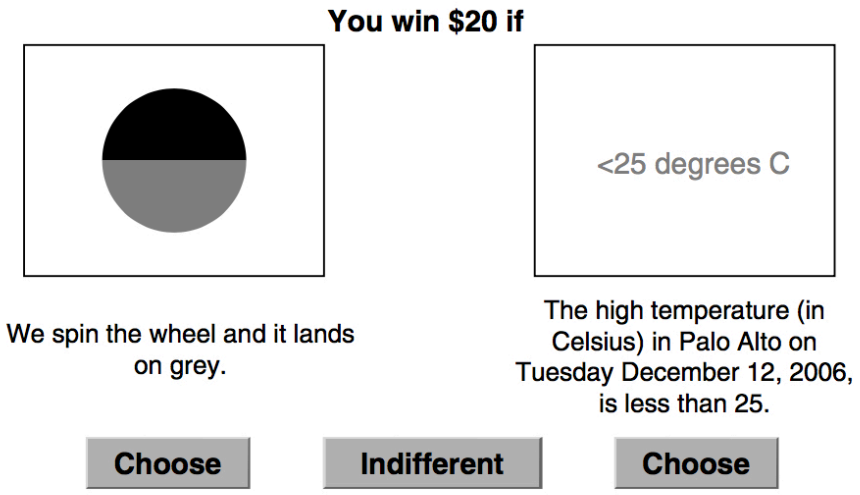

Figure 4.5: Elicitation screen proposed by Abbas et al. (2008) [1]

feedback on the information provided by the expert should be given during the elicitation process [368, 369, 361]. It is recommended that this feedback is presented in a graphical form [359, 279]. Ideally, it should also allow the expert to modify the information he provided so far in cases where he considers that it is after all not in accordance with his opinion.

4.4 The Bayesian hierarchical approach to account for exposure uncertainty

The most convenient way to construct complex probability models is via a hierarchical approach [370]. In the following, we will present the Bayesian hierarchical focusing on the specific aim to account for measurement error in an epidemiological study. Hierarchical models, which are often used to describe complex phenomena, allow to distinguish the different levels of information in a model [371]. More precisely, hierarchical models are composed of sub-models that are linked via conditional independence assumptions [372, 373, 370].

As described in chapter 3, in epidemiological studies, the data we observe are almost always imperfect and contaminated by missingness and mismeasurement. While this is generally true for both the exposure X and the outcome Y, the effects of the mismeasurement of the former will in general be more severe than the effects of the mismeasurement of the latter [6, 8]. As in chapter 3, we will therefore limit our discussion to exposure measurement error. In the situation, where instead of true exposure X, we merely dispose of observed exposure Z, which is contaminated by measurement error, it is no longer possible to describe the occurrence of the disease outcome Y by a probability model that is as simple as the model presented in section 4.2.1. Figure 4.6 illustrates how the different levels of the hierarchical model can be represented for the case of Berkson error. Following the terminology introduced by Clayton (1992) [374], and subsequently employed by Richardson [193, 194, 372, 9] in a series of papers treating measurement error in epidemiological studies, the probability model $[Y|\theta, X]$ is now denoted as disease model. It describes the probability of the occurrence of the disease outcome Y given true exposure X. In the hierarchical structure presented in Figure 4.6, we can account for the

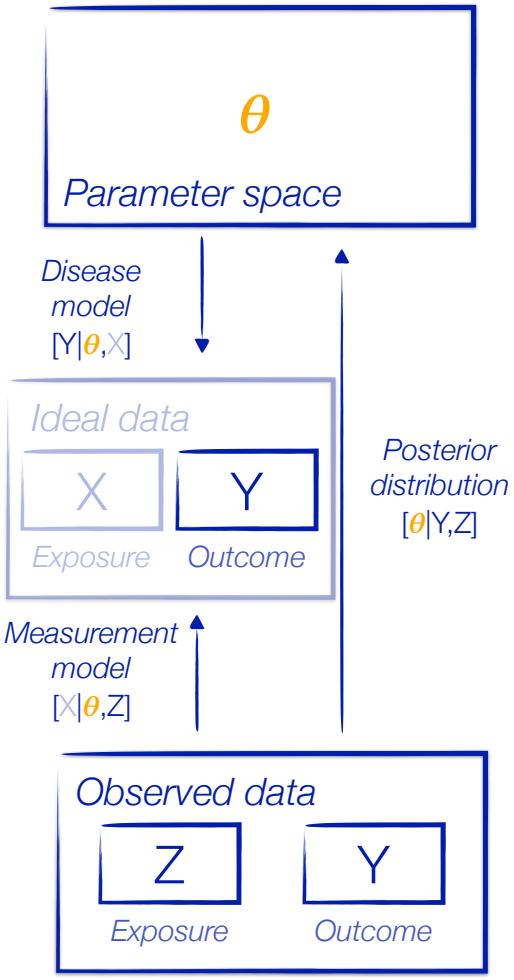

Figure 4.6: A sketch of the Bayesian hierarchical approach to account for Berkson error in an epidemiological study.

fact that we do not observe the values of X by treating true exposure as a latent variable, i.e. a variable that is not observed directly but only informed through observed exposure Z. Thereby, we can consider true exposure as part of an ideal data set, which we could have observed if we could measure exposure X without measurement error. The relationship between true X and observed Z exposure is described by the measurement model $[X|\boldsymbol{\theta}, Z]$. It is interesting to note that the treatment of measurement error can be seen as part of the more general framework of incomplete data problems, which also encompasses the treatment of missing and censored data [9, 6, 190].

The most natural way to represent the conditional independence assumptions of a hierarchical model is arguably via a Directed Acyclic Graph (DAG) [375]. Figure 4.7 shows a DAG corresponding to Berkson error. In this DAG, the vector of unknown parameters is split in two sub-vectors $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, all unknown quantities are represented in circles and observed quantities are represented in boxes. Arrows are used to indicate dependencies between the different quantities. This DAG illustrates how we can link the disease model, i.e. the conditional distribution of the outcome given true exposure and $\boldsymbol{\theta}_1$, $[Y|X, \boldsymbol{\theta}_1]$ and the measurement model, i.e. the conditional distribution of true exposure X given observed exposure Z, $[X|Z, \boldsymbol{\theta}_2]$. Additionally, the DAG highlights certain conditional independence assumptions, for instance the

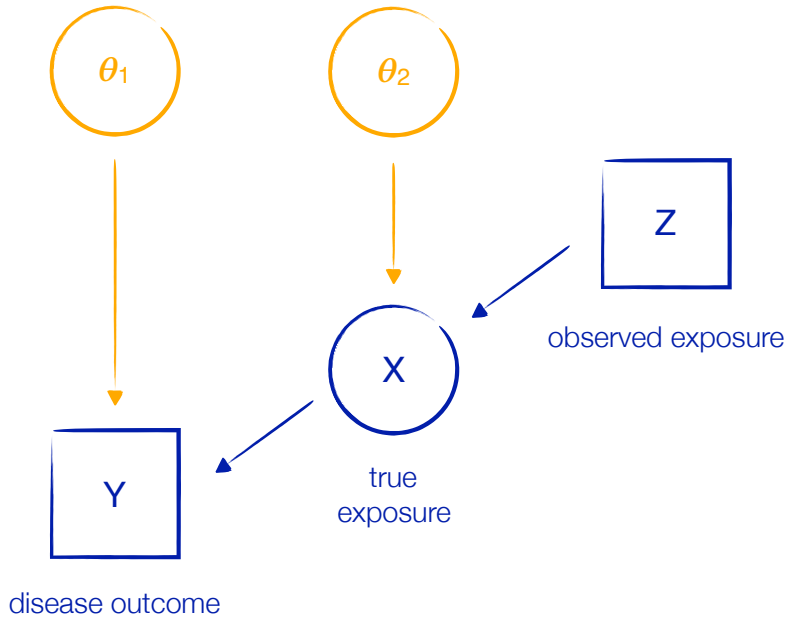

Figure 4.7: Directed Acyclic Graph (DAG) to describe Berkson error arising in an epidemiological study.

conditional independence of Y and Z given X ($[Y|X, Z, \boldsymbol{\theta}_1] = [Y|X, \boldsymbol{\theta}_1]$), which we described as the condition for non-differential measurement error in section 3.1.1.

In the case of classical measurement error, which is illustrated in the DAG presented in Figure

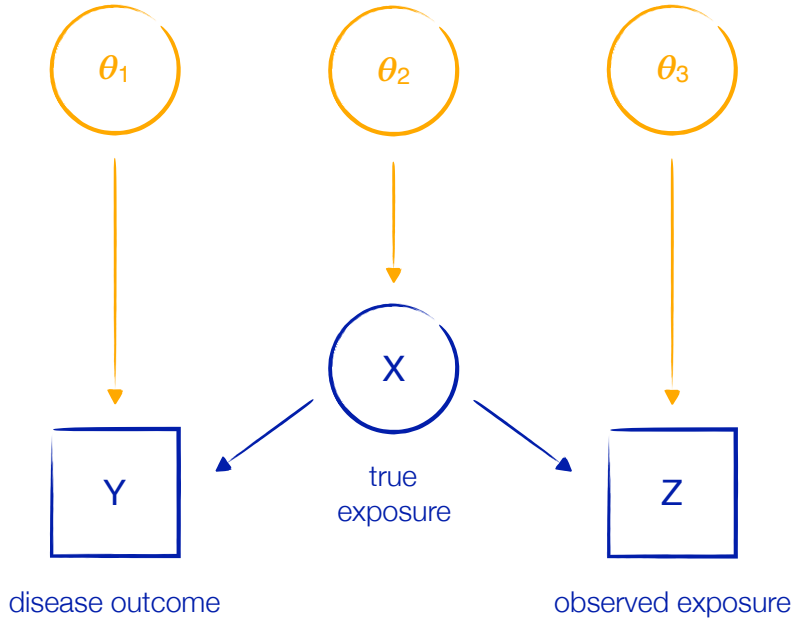

Figure 4.8: Directed Acyclic Graph (DAG) to describe classical measurement error arising in an epidemiological study.

4.8, the situation is slightly more complex. Similarly to Berkson error, we have to specify a disease model $[Y|X, \boldsymbol{\theta}_1]$ and a measurement model. However, contrary to Berkson error, where observed exposure Z is considered as a fixed covariate, for classical measurement error observed

exposure Z is now considered to be a random variable (see section 3.1.1). In this case, the measurement model describes the conditional distribution of observed exposure Z given true exposure X, i.e. $[Z|X, \boldsymbol{\theta}_3]$. To complete the specification of the hierarchical model, we need a third sub-model. This third sub-model, which is called exposure model, describes the distribution of unknown true exposure X as a function of the parameter vector $\boldsymbol{\theta}_2$. i.e. $[X|\boldsymbol{\theta}_2]$[9].
In the Bayesian hierarchical framework, it is not straightforward to derive the posterior distribution $[\boldsymbol{\theta}|Y, Z]$, i.e., the conditional distribution of the unknown parameters $\boldsymbol{\theta}$ given the observed data Y and Z. In the case of classical measurement error, this posterior distribution is defined as the following potentially high-dimensional integral:

$$[\boldsymbol{\theta}|Y, Z] = \int_X [\boldsymbol{\theta}, X|Y, Z]dX \tag{4.4}$$

$$\propto \int_X [\boldsymbol{\theta}, X, Y, Z]dX. \tag{4.5}$$

This integral often does not lead to an analytical expression of the posterior distribution. In this case, MCMC methods (presented in section 4.2.3) can be employed to generate samples from the posterior distribution and to marginalise over the values of the latent variable X [376], even though the sampling from this high-dimensional posterior distribution can be very time-consuming. Using the information in the DAG (Figure 4.8), we can rewrite the joint distribution for classical measurement error $[\boldsymbol{\theta}, X, Y, Z]$ in the following way:

$$[\boldsymbol{\theta}, X, Y, Z] = [Y|X, \boldsymbol{\theta_1}][Z|X, \boldsymbol{\theta}_3][X|\boldsymbol{\theta}_2][\boldsymbol{\theta}_1][\boldsymbol{\theta}_2][\boldsymbol{\theta}_3]. \tag{4.6}$$

Similarly to classical measurement error, we can rely on the assumption of non-differential error, to express the joint distribution $[\boldsymbol{\theta}, X, Y|Z]$ in the case of Berkson error as:

$$[\boldsymbol{\theta}, X, Y|Z] = [Y|X, \boldsymbol{\theta_1}][X|Z, \boldsymbol{\theta}_2][\boldsymbol{\theta}_1][\boldsymbol{\theta}_2]. \tag{4.7}$$

In contrast to classical functional approaches like regression calibration, where disjoint steps are used to estimate true exposure and unknown risk parameters, the Bayesian hierarchical approach allows to jointly estimate true exposure and all unknown parameters in a unique, global and coherent framework. Moreover, the Bayesian hierarchical approach to statistical inference is very flexible [9, 377, 203, 191], making it particularly suitable when the aim is to account for different sources of uncertainty. Indeed, in a structure resulting from the conditional linking of sub-models, we can complexify one of the sub-models without having to modify the other parts of the model and without compromising the validity of statistical inference. In particular, we can also introduce new levels in the hierarchical model, for instance a dose model if the aim is to simultaneously account for exposure and dose uncertainty.

[9] In the terminology introduced in section 3.2., the Bayesian hierarchical approach to account for measurement error is therefore a structural approach for measurement error correction.

Chapter 5

Methods

5.1 Presentation of the hierarchical models

We built several hierarchical models to account for exposure and dose uncertainty when modelling the association of radon exposure and lung cancer mortality in the French cohort of uranium miners. As described in section 4.4, this approach is based on the combination of sub-models, which are linked via conditional independence assumptions. When accounting for measurement error in radon exposure in the French cohort of uranium miners, the starting point is a disease model, which relates the right-censored variable time until death by lung cancer Y_i of miner $i, i \in \{1, \ldots, n\}$ to his true radon exposure $X_i(t)$ at time t and to p covariates $V_{i1}(t), \ldots V_{ip}(t)$. In the case of unshared classical measurement error, we can combine this disease model with a measurement model describing the conditional distribution of observed radon exposure of miner i at time t, $Z_i(t)$, given his true exposure $X_i(t)$, and with an exposure model, which specifies the distribution of true radon exposure $X_i(t)$. For Berkson error, on the other hand, we merely need to combine the disease model with a measurement model describing the conditional distribution of true radon exposure of miner i at time t, $X_i(t)$, given his observed exposure $Z_i(t)$ (see section 3.1.1 for the definitions of Berkson and classical measurement error). Owing to the flexibility of the Bayesian hierarchical approach, we can choose to complexify one of the sub-models without having to change the methods we employ to conduct statistical inference. In this vein, we can combine the disease model with different measurement models, which describe different types and magnitudes of error according to period of exposure in order to reflect changes in the methods of exposure assessment. Additionally, we can define measurement models which present a hierarchical structure themselves to describe more complex error structures, potentially including different components of shared or unshared measurement error (see section 3.1.1).

In order to account for dose uncertainty, on the other hand, the disease model can be combined with a dose model, which calculates absorbed lung dose as a function of the most important input parameters described in section 2.3.3.

Finally, it is possible to combine the disease model, the measurement model, the exposure model and the dose model in order to account for both exposure and dose uncertainty in a unique and coherent inferential framework when estimating the parameters of interest. In this section, we will present the different sub-models we assumed in this work, describe how to combine them in

a hierarchical structure and specify the chosen prior distributions which are necessary to conduct Bayesian inference for the full hierarchical model.

5.1.1 The disease model

When modelling lung cancer mortality in the French cohort of uranium miners, the outcome of interest is the time until death by lung cancer of miner i ($i = 1, ..., n$), where $n = 5086$ is the total number of uranium miners in the cohort. As not all workers in the French cohort of uranium miners died of lung cancer, this outcome is a censored variable. To be more precise, it is right-censored, being defined as the earliest among time until death by lung cancer of miner i, T_i and a censoring variable C_i. The censoring variable C_i is again defined as the earliest among time until death by a cause other than lung cancer, time until loss to follow-up, time until the 31^{st} December 2007 and time until the 85^{th} birthday (see section 2.2.2). The observed outcome can therefore be represented by the n pairs of random variables (Y_i, δ_i), where $Y_i = \min(T_i, C_i)$ is a non-negative continuous variable and δ_i is a binary variable with $\delta_i = 1$ if miner i died of lung cancer at time Y_i (i.e., $T_i \leq C_i$) and $\delta_i = 0$ if he "would have died of lung cancer" after time C_i (i.e., $T_i > C_i$).
As mentioned in section 2.2.4, when conducting statistical inference on the association between radon exposure and lung cancer mortality in cohorts of uranium miners, the most common approach is to adopt a grouped Poisson regression model. In theory, this modelling approach, which is also known as the life table approach to survival analysis [378] is equivalent to the analysis of survival times via proportional hazards models in the case of a piecewise exponential survival distribution and categorical covariates [379]. This approach has many advantages over the classical choice to analyse survival times via proportional hazards models including greater flexibility in the modelling of multiple events and the fact that it allows to analyse survival times in the context of a generalised linear model. However, when analysing the association between radon exposure and lung cancer mortality it involves the categorisation of radon exposure and the fact that all information has to be averaged over the person-years in a stratum may imply a substantial loss of information on the variance in each stratum. Moreover, we have seen in section 3.1.2 that the categorisation of a continuous exposure variable which is prone to non-differential measurement error can result in differential misclassification. Finally, a stratified Poisson regression model impedes the modelling of measurement error at the individual level, which is often its natural level of occurrence [136]. Based on these arguments, we chose to model the individual survival and failure times of the 5086 miners in the French cohort of uranium miners via a proportional hazards model. In this context, we model the instantaneous hazard rate of death by lung cancer of miner i at time t, $h_i(t)$. If $f_i(t)$ and $F_i(t)$ denote the probability density function and the cumulated density function of T_i, respectively, the instantaneous hazard rate is defined as:

$$h_i(t) = \lim_{\Delta t \to 0} \frac{P(t < T_i \leq t + \Delta t | T_i > t)}{\Delta t} = \frac{f_i(t)}{1 - F_i(t)} = \frac{f_i(t)}{S_i(t)} \tag{5.1}$$

where $S_i(t) = 1 - F_i(t) = \int_t^\infty f_i(u)du$ is called the survival function of miner i at time t. $\Delta t \cdot h_i(t)$, is the approximate probability for miner i to die of lung cancer in the time interval $[t, t + \Delta t]$ given that he survived until time t [380]. Based on (5.1), we can easily express $f_i(t)$

as $f_i(t) = h_i(t) \cdot S_i(t)$. Moreover, given that $f_i(t) = \frac{d}{dt}F_i(t) = -\frac{d}{dt}(1 - F_i(t)) = -\frac{d}{dt}S_i(t)$, we can derive

$$h_i(t) = \frac{-\frac{d}{dt}S_i(t)}{S_i(t)} = -\frac{d}{dt}\log S_i(t). \tag{5.2}$$

Finally, this result leads to the following expression for the survival function $S_i(t)$ and the probability density function $f_i(t)$:

$$S_i(t) = \exp\left(-\int_0^t h_i(u)du\right) \tag{5.3}$$

and

$$f_i(t) = h_i(t)\exp\left(-\int_0^t h_i(u)du\right). \tag{5.4}$$

If we denote $g_i(c)$ and $G_i(c) = P(C_i > c)$ the probability density and the survival function for the censoring variable C_i, and under the assumption that C_i and the failure time variable T_i are independent, the contribution of the pair of observed random variables $(Y_i, \boldsymbol{\delta}_i)$ to the likelihood is given by:

$$[Y_i = y_i, \boldsymbol{\delta}_i = \delta_i] = [T_i = y_i; C_i > y_i]^{\delta_i}[C_i = y_i; T_i > y_i]^{1-\delta_i} \tag{5.5}$$

$$= f_i(y_i)^{\delta_i}G_i(y_i)^{\delta_i}g_i(y_i)^{1-\delta_i}S_i(y_i)^{1-\delta_i}. \tag{5.6}$$

In order to avoid the specification of $g_i(c)$ and $G_i(c)$, it is common in survival analysis to make the assumption of non-informative censoring [381, 382], according to which the distribution of C_i provides no information about the distributions of T_i. In this case, we can treat both $G_i(c)$ and $g_i(c)$ as constants and consider

$$[Y_i, \delta_i] \propto f_i(y_i)^{\delta_i}S_i(y_i)^{1-\delta_i} = h_i(y_i)^{\delta_i}S_i(y_i). \tag{5.7}$$

The most popular approach to analyse survival data is to assume a proportional hazards model [383]. In the case of the French cohort of uranium miners, we can assume the following semi-parameteric model:

$$h_i(t;\boldsymbol{\theta}) = h_0(t)\varphi(X_i^{\text{cum}}(t), V_{i1}(t), ..., V_{ip}(t);\boldsymbol{\theta}), \tag{5.8}$$

where $\boldsymbol{\theta}$ and the so-called hazard ratio $\varphi(X_i^{\text{cum}}(t), V_{i1}(t), ..., V_{ip}(t);\boldsymbol{\theta})$ is multiplied by the baseline hazard $h_0(t)$ which is assumed to be the same for all miners. The hazard ratio is a positive term expressing how the instantaneous hazard rate varies as a function of cumulated exposure to radon of miner i at time t, $X_i^{\text{cum}}(t)$ (expressed in 100 WLM) and of a number of potential effect modifying variables $V_{i1}(t), \ldots, V_{ip}(t)$, that will be specified in the next section.
Cumulated exposure $X_i^{\text{cum}}(t)$ was lagged by five years in order to allow for a latency period between a received exposure value and its potential impact on health outcomes [384, 385, 386], thereby excluding exposure values that were received immediately before the death by lung cancer. This exposure lag implies that the instantaneous hazard rate of miner i at time t, $h_i(t)$, is not modelled as a function of the cumulated exposure received until time t, $X_i^{\text{cum}}(t)$, but as a function of the cumulated exposure received until five years before time t. Note that this exposure lag would unduly complicate the notation of the proportional hazards models presented in this section. For ease of readability, this exposure lag of five years will therefore not appear in

the following expressions. Concerning the choice of time-scale in proportional hazards models, time-on-study and age are the most common options. Inasmuch as we model occupational exposure, the use of time-on-study as time-scale may seem to be a plausible choice, but Kleinbaum [381] suggests to favour age as time-scale whenever age at event is likely to have a larger effect on the hazard than time-on-study. Based on previous findings on cohorts of uranium miners, we can assume that, contrary to the attained age of a miner [83], the timing of study initiation has no inherent meaning in terms of the risk of lung cancer mortality in the cohort. Furthermore, several authors recommend to favour age as time-scale whenever possible [387, 388, 389] since the modelling of the effect of age can be complex and prone to misspecification errors. Based on these arguments, we chose attained age as time scale. This choice leads to a left-truncation of both T_i and C_i [390], implying the same left-truncation for the minimum of these two variables, Y_i. Indeed, all variables must take values greater than the age at study entry into the cohort. In other words, $P(T_i < r_i^0) = 0$, $P(C_i < r_i^0) = 0$ and $P(Y_i < r_i^0) = 0$, where r_i^0 is the age of miner i at entry into the French cohort of uranium miners. Additionally, the choice to censor workers at their 85^{th} birthday implies that C_i is not only left-, but also right-truncated, thereby resulting in a right-truncation of Y_i, whereas T_i can of course take values greater than 85 years. As we merely disposed of information on the annual radon exposure of each miner, it may seem reasonable to choose a rather coarse modelling of survival times by choosing a time-scale of age in years. As this choice may imply a substantial loss of information, we preferred to model age in days. In this context, we had to choose a date on which this annual exposure was received every year. To avoid the overestimation of cumulated radon exposure of a miner at a given time, we chose the 31^{st} December for this purpose. Thereby, the time-varying pattern of cumulated radon exposure of miner i at time t was assumed to follow a step function with jumps on the 31^{st} December of each year a miner received an exposure. While it may appear more reasonable to interpolate values or to assume smaller intervals of time, Therneau et al. (2000) discuss this point and conclude that these refinements make little difference in practice [391]. In contrast, it is very important to properly account for the nature of time-varying variables in proportional hazard models by linking the full exposure history to the instantaneous hazard rate. In this vein, it would induce a bias in risk estimates if we considered the total cumulated radon exposure a miner received during his career to be known at study entry, i.e. if we pretended to be able to look into the future [392]. This bias, which is common in the medical literature [393, 394, 390], is classically known as time-dependent bias [395]. It occurs whenever the future exposure status of a study participant is analysed as being known at the beginning of follow-up [394].

Both the left-truncation of T_i and the time-varying nature of cumulated exposure can be accounted for via the survival function $S_i(t;\boldsymbol{\theta})$. In the case of a left-truncation at time r_i^0 as defined before, the survival of miner i until time t can be expressed as

$$\exp\left(-\int_{r_i^0}^{t} h_i(u;\boldsymbol{\theta})du\right) = \exp\left(-\int_{0}^{t} h_i(u;\boldsymbol{\theta})du + \int_{0}^{r_i^0} h_i(u;\boldsymbol{\theta})du\right) = \frac{\exp\left(-\int_0^t h_i(u;\boldsymbol{\theta})du\right)}{\exp\left(-\int_0^{r_i^0} h_i(u;\boldsymbol{\theta})du\right)} = \frac{S_i(t;\boldsymbol{\theta})}{S_i(r_i^0;\boldsymbol{\theta})}. \tag{5.9}$$

To account for the time-varying pattern in cumulated exposure, we only have to apply this reasoning for the different time periods during which cumulated radon exposure is constant for miner i. Let $r_i^0 < r_i^1 < \ldots < r_i^{M_i} = y_i$ be a finite partition into M_i intervals of the time interval $(r_i^0, y_i]$. Let $\omega_{i,m}$ be the constant value of the cumulated exposure to radon of miner i on the

m^{th} time-interval $(r_i^{m-1}, r_i^m]$ and $S_i^{\omega_{i,m}}$ denote the survival function of miner i with a cumulated radon exposure of $\omega_{i,m}$. In this context, following the same reasoning as (5.9), the survival function of miner i at time y_i can be expressed as:

$$\begin{aligned} S_i(y_i;\boldsymbol{\theta}) &= \exp\left(-\int_{r_i^0}^{y_i} h_i(u;\boldsymbol{\theta})du\right) \\ &= \prod_{m=1}^{M_i} \exp\left(-\int_{r_i^{m-1}}^{r_i^m} h_i(u;\boldsymbol{\theta})du\right) \\ &= \prod_{m=1}^{M_i} \frac{S_i^{\omega_{i,m}}(r_i^m;\boldsymbol{\theta})}{S_i^{\omega_{i,(m-1)}}(r_i^{m-1};\boldsymbol{\theta})} \end{aligned} \tag{5.10}$$

In line with this reasoning, statistical inference in proportional hazards models in the presence of time-varying covariates is commonly based on so-called "pseudo subjects" where each period of a subject which is associated with a constant value of a time-varying covariate is represented as an independent row (which represents a pseudo subject) in the data set [391, 392]. In the French cohort of uranium miners, the expression of the time-varying exposure history of miners via pseudo miners resulted in a data frame with 58447 rows for the 5086 miners in the full cohort and in 40162 rows for the 3377 miners in the post-55 cohort.

Modelling the hazard ratio

When describing the hazard ratio $\varphi(X_i^{\text{cum}}(t), V_{i1}(t), ..., V_{in}(t), \boldsymbol{\theta})$, we considered the two model structures

$$\varphi(X_i^{\text{cum}}(t), V_{i1}(t), ..., V_{in}(t), \boldsymbol{\theta}) = 1 + \beta X_i^{\text{cum}}(t) \tag{5.11}$$

$$\varphi(X_i^{\text{cum}}(t), V_{i1}(t), ..., V_{in}(t), \boldsymbol{\theta}) = \exp\left(\beta \cdot X_i^{\text{cum}}(t)\right), \tag{5.12}$$

where β is an unknown parameter that quantifies the risk of lung cancer mortality associated with cumulated radon exposure $X_i^{\text{cum}}(t)$. The first model (5.11), referred to as Excess Hazard Ratio (EHR) model in the following, presents a linear structure, which is commonly used to describe the association between solid cancer mortality and exposure to radon and to other sources of ionising radiation. This model will be denoted as $\mathcal{D}_1$ or as linear EHR model in the following. The second model (5.12), is a log-linear model and presents the more classical form of proportional hazards model proposed by Cox (1972) [180]. It will be referred to as $\mathcal{D}_2$ or the Cox model in the following. Note that the Cox model inherently respects the condition that the instantaneous hazard rate $h_i(t;\boldsymbol{\theta}) = h_0(t)\ \varphi(X_i^{\text{cum}}(t), V_{i1}(t), ..., V_{in}(t), \boldsymbol{\theta})$ must be positive, while for the EHR model the parameter values are subject to the constraint $\beta X_i^{\text{cum}}(t) > -1$ $\forall t \forall i$.

Apart from model $\mathcal{D}_1$, we considered several other structures for the excess hazard ratio to account for potential effect modifying variables. Vacquier et al. (2009) [70] identified period of exposure (until 1955 and after 1955) as the most important effect modifying variable in the French cohort of uranium miners. The corresponding model ($\mathcal{D}_3$) for the excess hazard ratio is $\varphi(X_i^{\text{cum}}(t), V_{i1}(t), ..., V_{in}(t), \boldsymbol{\theta}) = 1 + \beta_1 \cdot X_{i,\leq 1955}^{\text{cum}}(t) + \beta_2 \cdot X_{i,>1955}^{\text{cum}}(t)$ with $X_{i,\leq 1955}^{\text{cum}}(t)$ and $X_{i,>1955}^{\text{cum}}(t)$ the cumulated radon exposure until time t that miner i received until 1955 and after

1955, respectively. The associated effect modifying variable $V_{1i}(t) = \mathbb{1}_{>1955}(t)$ indicates whether an exposure for miner i at time t was received after 1955 or not. $X^{\text{cum}}_{i,>1955}(t)$ can be expressed through $V_{1i}(t)$ by $X^{\text{cum}}_{i,>1955}(t) = \sum_{l=r_i^0}^{t} V_{1i}(l)X_i(l)$, where $X_i(l)$ is the radon exposure a miner i received at time l. Likewise, $X^{\text{cum}}_{i,\leq 1955}(t)$ can be expressed by $X^{\text{cum}}_{i,\leq 1955}(t) = \sum_{l=r_i^0}^{t}(1-V_{1i}(l))X_i(l)$. Note that the exposures of a miner have to be summed over all times l at which a miner received a radon exposure before time t, $l = r_i^0, ..., t$.

We also chose to include time since exposure as modifying variable ($\mathcal{D}_4$). In this context, the hazard ratio associated with radon exposure is modelled as $\varphi(X_i^{\text{cum}}(t), V_{i1}(t), ..., V_{in}(t), \boldsymbol{\theta}) = 1 + \beta_1 X^{\text{cum}}_{i,5-14}(t) + \beta_2 X^{\text{cum}}_{i,15-24}(t) + \beta_3 X^{\text{cum}}_{i,25+}(t)$, where $X^{\text{cum}}_{i,5-14}(t), X^{\text{cum}}_{i,15-24}(t)$ and $X^{\text{cum}}_{i,25+}(t)$ are the cumulated radon exposure at time t that miner i received in the last 15 years, 15 to 25 years ago or more than 25 years ago, respectively. $X^{\text{cum}}_{i,5-14}(t)$ does not include exposures received in the last 5 years to respect the exposure lag of five years defined earlier. In the same way as for period of exposure, two indicator variables $V_{2i}(t)$ and $V_{3i}(t)$ can be defined to express the effect modification by time since exposure. Note that when considering time since exposure as effect modifying variable, we have to account for all changes in the time-varying variables $X^{\text{cum}}_{i,5-14}(t), X^{\text{cum}}_{i,15-24}(t)$ and $X^{\text{cum}}_{i,25+}(t)$, for instance by considering the different exposure periods for which all three exposure variables were constant as pseudo miners as described earlier. In the French cohort of uranium miners, accounting for the time-varying nature of these three exposure variables via pseudo miners resulted in a data frame with 119073 rows for the 5086 miners in the full cohort and in 77919 rows for the 3377 miners in the post-55 cohort. As we chose attained age as timescale, the variables age at exposure and time since exposure contain essentially the same information [72]. Finally, we compared these models with two piecewise linear EHR models with breakpoints at 50 WLM ($\mathcal{D}_5$) and 100 WLM ($\mathcal{D}_6$) in order to assess the linearity of the EHR model $\mathcal{D}_1$. Depending on the disease model, the vector of unknown risk coefficients $\boldsymbol{\beta}$ therefore either consisted only of β ($\mathcal{D}_1$ and $\mathcal{D}_2$), of β_1 and β_2 ($\mathcal{D}_3$, $\mathcal{D}_5$ and $\mathcal{D}_6$) or of β_1, β_2 and β_3 ($\mathcal{D}_4$).

Modelling the baseline hazard

When conducting frequentist inference on the unknown parameter in a proportional hazards model, the baseline hazard $h_0(t)$ is usually unspecified and inference is based on the partial instead of on the full likelihood [382, 381, 183]. As described in section 3.1.2, this approach no longer yields valid statistical inference on risk parameters in the presence of measurement error [181, 182, 183]. A convenient solution in this situation is to specify a parametric model for the baseline hazard $h_0(t; \boldsymbol{\theta})$. In the context of frequentist inference it is difficult to simultaneously estimate the parameters describing the baseline hazard and the parameters describing the hazard ratio [383, 396]. While it is therefore common to obtain the parameters of the baseline hazard via plug-in estimates in the frequentist approach to statistical inference, the Bayesian approach allows to simultaneously estimate all unknown parameters in a coherent framework [383, 396].

When specifying the baseline hazard, one of the most convenient and popular choices is to use a semi-parametric model where the baseline is supposed to be piecewise-constant [397, 380, 383, 149, 398, 396, 399], as illustrated in Figure 5.1. Given a finite partition of the time axis

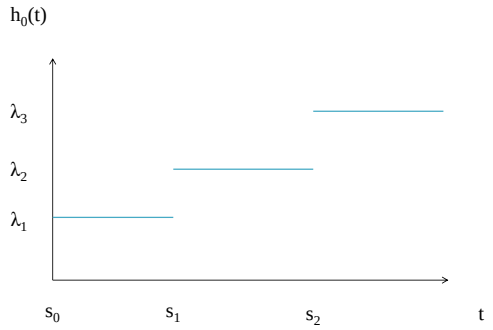

Figure 5.1: Example of a piecewise constant model for the baseline hazard in a proportional hazards model

$0 = s_0 < s_1 < s_2 < \ldots < s_J$ with $s_J = \max_{i=1,\ldots,n} Y_i$ we define the baseline hazard $h_0(t;\boldsymbol{\theta})$ as

$$h_0(t;\boldsymbol{\theta}) = \lambda_j \qquad \forall t \in I_j = (s_{j-1}, s_j] \tag{5.13}$$

where I_j denotes the j^{th} time interval ($j = 1, ..., J$). When assuming a piecewise-constant model on baseline hazard and a linear EHR model on the hazard ratio (i.e. model $\mathcal{D}_1$), the survival function at time t for a miner i with constant cumulated radon exposure ω and without left-truncation can be written as:

$$S_i(t;\boldsymbol{\theta}) = \exp\left(-(1+\beta\omega)\sum_{j=1}^{J}\nu_{ij}\left(\lambda_j(t-s_{j-1}) + \sum_{g=1}^{j-1}\lambda_g(s_g - s_{g-1})\right)\right) \tag{5.14}$$

where $\boldsymbol{\theta} = (\lambda_1, \cdots, \lambda_J, \beta)$ and $\nu_{ij} = 1$ if $y_i \in (s_{j-1}, s_j]$ and 0 otherwise.
In the context of this model, one has to choose cut-points $s_1, \ldots, s_J$ to define the intervals for which the values of the baseline hazard are constant. Heidenreich et al. [36], who analysed lung cancer mortality in a large cohort of uranium miners, chose nine age intervals for this purpose, i.e., one for the baseline hazard before 40 years of age, one for the baseline hazard after the age of 70 and intervals of five years in between these two boundaries. We used the same lower boundary of 40 years of age, but decided for a more parsimonious model with steps of 15 years, given the small number of miners deceased by lung cancer in our study compared to Heidenreich et al. [36] (i.e. 211 vs. 1538). The cut-points of the time-axis s_1, s_2, s_3, s_4 we chose were thus 40, 55, 70 and 85 years.

We used data on the general French male population [400] in order to investigate the way in which the baseline hazard of lung cancer mortality might have changed over time. Figure 5.2 illustrates the time trend in the lung cancer mortality rate in French males for four different time periods. As general lung cancer mortality rate changed substantially in the time period of our follow-up, i.e. between 1946 and 2007 [401], it might seem natural to model the baseline hazard not only as a function of attained age, but also of calendar period. Figure 5.2 shows marked difference between the first time period (1968-1977 in red in figure 5.2) and the three subsequent time periods for the age groups over 70 years. There are only two subjects in the cohort of French uranium miners who died of lung cancer at an age over 70 and before 1978.

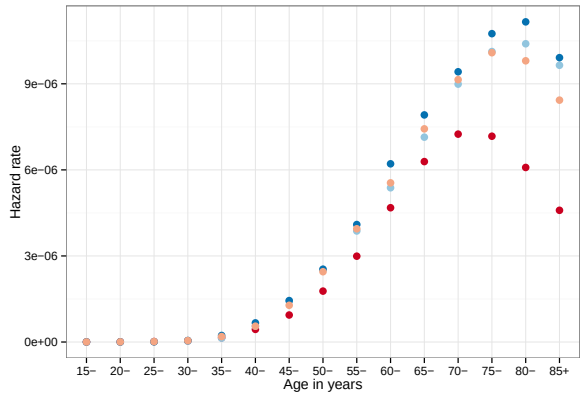

Figure 5.2: Lung cancer mortality rate in French males for the periods 1968-1977 (red), 1978-1987 (orange), 1988-1997 (blue), 1998-2005 (lightblue)

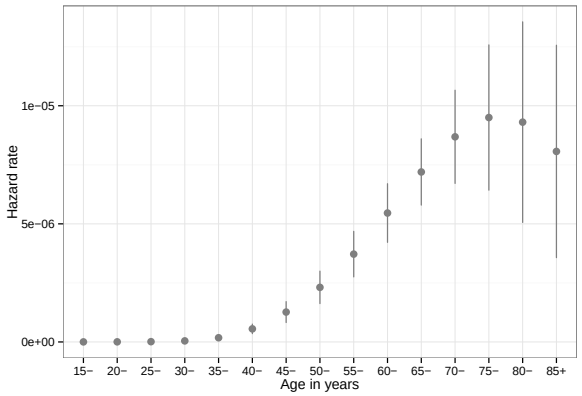

Figure 5.3: Average hazard rate of lung cancer mortality in French males for the period 1968-2005 with associated 95% intervals of values for this time period

For the sake of parsimony, we thus decided to model baseline hazard solely as a function of age and not of calendar period. Figure 5.3 summarises the empirical distribution of the mortality

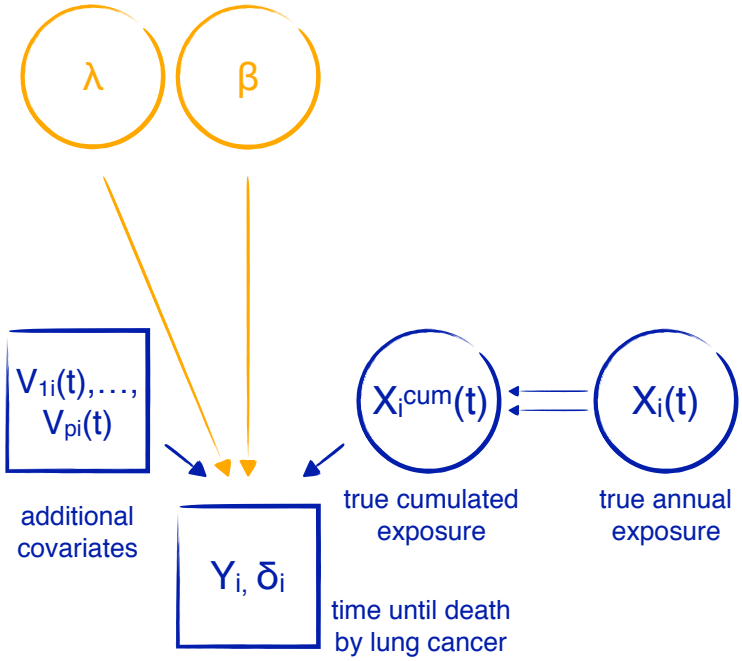

Figure 5.4: Directed acyclic graph of the disease model. Circles indicate unknown quantities and rectangles indicate observed variables. Single arrows indicate a probabilistic dependency while double arrows indicate a deterministic dependency. $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ are the vector of baseline parameters and the vector of risk parameters intervening in the baseline hazard and the hazard ratio, respectively.

hazards calculated on all available values in the time period 1968-2005. We used this external data to specify an informative prior probability distribution on the unknown parameters $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ as will be described in section 5.1.6. The probabilistic dependencies between the variables in the disease model are represented in the Directed Acyclic Graph (DAG) in Fig-

ure 5.4.

5.1.2 Measurement models

In this section, we will present the different measurement models that we considered to describe the shared and unshared exposure uncertainty components that can be assumed for the French cohort of uranium miners. As described in section 2.2.3, the methods of exposure assessment in the cohort changed over time. Between 1946 and 1956, annual radon exposure values were reconstructed retrospectively by a group of experts. In this reconstruction, a monthly exposure value was estimated for each mining division and each year of exposure between 1946 and 1956. These estimates were based on environmental measurements performed in the mines and on information concerning the working conditions in the different mining divisions. In 1956, measurements of ambient radon gas concentration at work sites were introduced for the purpose of radiation protection. These ambient measurements of radon exposure allowed for a prospective method of exposure assessment in the mines. Finally, in 1983, personal dosimetry was introduced in the mines. Based on these changes in the methods of exposure assessment, measurement models describing the exposure uncertainty characteristics in the cohort have to account for potential changes in the nature and magnitude of exposure measurement error.
In all measurement models presented in this section, we will denote $X_i^q(t)$ and $Z_i^q(t)$ the true and the observed exposure of miner i at time t in the q^{th} exposure period, respectively. When accounting for unshared measurement error, we assumed five different exposure periods following the characterisation of measurement error in the French cohort of uranium miners made by Allodji et al. (2012) [33, 34]. Thereby, $q \in \{1, 2, 3, 4, 5\}$ corresponds to the exposure periods 1946-1955, 1956-1974, 1975-1977, 1978-1982 and 1983-2007, respectively. When modelling shared measurement error, on the other hand, our aim was to account for the fact that there were components of error that might be shared for a miner for the exposure years between 1956 and 1982, as they all relied on a group-exposure assessment strategy based on the measurements of ambient radon gas concentrations. We therefore considered the exposure years between 1956 and 1982 as a homogeneous exposure period when accounting for shared exposure measurement error. In this situation, we assumed $q \in \{1, 2, 3\}$ corresponding to the exposure periods 1946-1955, 1956-1982 and 1983-2007.
A large part of the measurement error literature is based on additive error. It has been repeatedly suggested, however, that a multiplicative measurement error model may be more realistic in occupational and environmental epidemiology in general [147] and to describe uncertainty in airborne exposure in particular [31, 152, 189]. In the following, we will therefore postulate a log-normal and multiplicative error structure to represent the measurement of radon exposure in the French cohort of uranium miners. This error structure guarantees the positivity of both true and observed exposure and the hypothesis of log-normal measurement errors is in accordance with much of the literature on exposure uncertainty in radon exposure [24, 25, 402, 28, 29, 30, 151, 40, 36, 33, 34, 35]. In line with this hypothesis, repeated measurements of radon concentrations provide evidence for a multiplicative structure of measurement error in radon exposure [26, 403]. Heid (2002) [402] gives a further justification of this error structure based on the fact that many factors modifying radon concentration affect a proportion

of radon atoms in a given environment. We will thus describe the distribution of the measurement error term $U_i^q(t)$ concerning miner i at time t during the exposure period q by specifying the mean and the variance of $\log\left(U_i^q(\,t)\right)$ as $-\frac{\sigma_{U,q}^2}{2}$ and $\sigma_{U,q}^2$, respectively. This parametrisation implies that $E\left(U_i^q(t)\right) = 1$ and thereby that the measurement error term does not introduce a systematic under- or overestimation of exposure.

Accounting for unshared measurement error

When characterising exposure uncertainty in the French cohort of uranium miners, Allodji et al. (2012) [33] identified the variation in airborne radon gas concentration as the most important of six sources of uncertainty associated with the measurement of radon exposure between 1956 and 1982. The remaining sources of uncertainty include the precision of the measurement device, human errors made by radiological protection operators, the estimation of working time, the estimation of a factor describing the chemical equilibrium between radon and its progeny and errors occurring during record-keeping and data transcription. Concerning the exposure years between 1956 and 1982, which were characterised by a prospective method of group level exposure assessment through ambient measurements, the authors identified three distinct exposure periods with varying magnitude of uncertainty associated with radon exposure measurement. The definition of these periods was based on factors including the number of measurements, the measurement device used as described above and the type of ventilation present in the mines. Moreover, they described the measurement errors occurring in these exposure periods, as well as the measurement errors occurring in the first exposure period, which was characterised by the exposure estimations by experts as Berkson error with $E(U_i^q(t)|Z_i^q(t)) = 1$, $\forall q \neq 5$ [34, 35]. This can be justified by the fact that in these four exposure periods, an estimate of radon exposure for a certain position in a mine was assigned to all miners present at a given place and time. The true exposure $X_i^q(t)$ of a miner i might differ from this assigned exposure value $Z_i^q(t)$, mainly due to variations in airborne radon concentration.
Exposure measurements produced by individual dosimetry in the fifth exposure period (1983-2007), on the other hand, are supposed to produce independent classical measurement error: $E(U_i^5(t)|X_i^5(t)) = 1$. Indeed, for this time period, it is plausible to assume that the true individual radon exposure of miner i is merely assessable with errors, which are essentially caused by a lack of precision of his individual measurement device, i.e., his personal dosimeter. In line with the assessment of exposure uncertainty in the cohort [33, 34, 35], we considered the following measurement model for the five exposure periods:

$$\boldsymbol{\mathcal{M}_1}: \begin{cases} X_i^1(t) = Z_i^1(t) \cdot U_i^1(t) \\ X_i^2(t) = Z_i^2(t) \cdot U_i^2(t) \\ X_i^3(t) = Z_i^3(t) \cdot U_i^3(t) \\ X_i^4(t) = Z_i^4(t) \cdot U_i^4(t) \\ Z_i^5(t) = X_i^5(t) \cdot U_i^5(t). \end{cases}$$

In other words, we assumed heteroscedastic and unshared measurement error to account for the fact that the type and the magnitude of error changed over the years.

When accounting for unshared measurement error in the cohort, we will thus suppose five different variance parameters $\sigma^2_{U,1}, \sigma^2_{U,2}, \sigma^2_{U,3}, \sigma^2_{U,4}, \sigma^2_{U,5}$ of $\log\left(U_i^q(t)\right)$ associated with the five distinct exposure periods identified by [33]. In this case, we will denote $\boldsymbol{\sigma}^2_U = (\sigma^2_{U,1}, \sigma^2_{U,2}, \sigma^2_{U,3}, \sigma^2_{U,4}, \sigma^2_{U,5})$ the vector of unknown variance parameters of $\log\left(U_i^q(t)\right)$.

Accounting for shared measurement error due to individual job conditions and worker practices

As described in section 3.1.1, measurement errors may be shared between individuals or for several years of exposure of the same individual. In particular, the methods of group exposure assessment which were employed in the French cohort of uranium miners before 1982 are likely to give rise to error components that are shared for several exposure years of the same miner. Indeed, when occupational exposure estimation is based on a strategy of group-exposure assessment, individual job conditions and worker practices can create a statistical dependence between the measurement errors in the exposure history of a worker [15, 16]. For instance, a comparison between exposures obtained through ambient measurements and through individual dosimetry in the French cohort of uranium miners revealed that some miners received considerably higher individual monthly exposure values than what would be expected through ambient measurements [78]. This discrepancy might be explained by the fact that some of the workers sought relief from the strong airstream produced by a ventilation system in their break hours, thereby exposing themselves to very high radon concentrations. As the areas where the airflow was weak were officially established as forbidden areas in the mines, no ambient measurements were taken in these areas. The habit of miner i to spent his break hours in these forbidden areas can therefore lead to an individual error component, which remains constant for several exposure years when exposure measurement is based on a group exposure assessment strategy, i.e. $U_i^q(t) = U_i^q \ \forall t$. Likewise, the specific job conditions of a miner may also influence this subject-specific error component U_i^q. If miner i habitually works in a specific place in the mine where ambient measurements are commonly taken, it is likely that his subject-specific error component U_i^q, which is common to all years of exposure in a given period q, is smaller than for a miner habitually working in a place in the mine where ambient measurements are more rare. Moreover, as mentioned earlier, Allodji et al. (2012) [33] identified the natural variations in airborne radon gas concentration as the most important source of exposure uncertainty before 1983. Several authors have shown that radon concentrations in homes and in mines present important diurnal and seasonal fluctuations [404]. As the work in the mines was organised by shifts, the habitual shift of a miner may influence his subject-specific measurement error component U_i^q. In the same way, a habit to take his holidays in a certain month of a year may influence U_i^q, at least for the exposure period before 1956 when monthly exposure values were estimated for a given mine and year, i.e. for $q = 1$. To describe the influence of these individual worker practices and job conditions, we proposed a second measurement model in which the Berkson error occurring in the period before 1983 is shared for several years of exposure of the same miner. In particular, we assumed a common error component U_i^1 for miner i for all years of retrospective exposure assessment and another common error component U_i^2 for miner i for the exposure years that were characterised by ambient measurements. Since exposure estimates after 1983 were based on personal dosimetry, there is no reason to assume a shared error com-

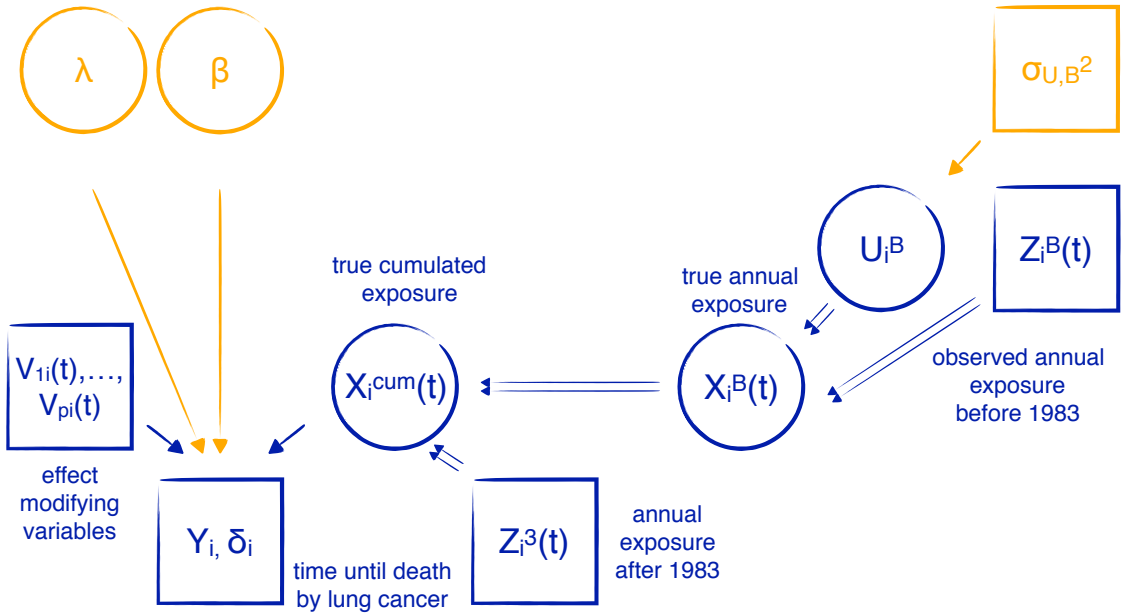

Figure 5.5: DAG of the full hierarchical model accounting for shared Berkson error until 1982 and neglecting unshared measurement error for the exposure period 1983-2007. B takes values 1 or 2.

ponent for this exposure period. Consequently, we only assumed three periods $q \in \{1, 2, 3\}$ when accounting for shared measurement error corresponding to the exposure years 1946-1955, 1956-1982 and 1983-2007. Moreover, as the magnitude of exposure measurement error which occurred during the last exposure period (1983-2007) can be considered to be much smaller than in earlier periods, we neglected the measurement error occurring in this last period when accounting for shared measurement error. The radon exposure observed between 1983 and 2007 may be considered to be the true exposure and we will merely focus on the modelling of the error components which occurred between 1946 and 1982.

The second measurement model that we assumed in this work can thereby be expressed as follows:

$$\boldsymbol{\mathcal{M}_2}: \begin{cases} X_i^1(t) = Z_i^1(t) \cdot U_i^1 \\ X_i^2(t) = Z_i^2(t) \cdot U_i^2 \\ Z_i^3(t) = X_i^3(t). \end{cases}$$

In this model, we assumed $\log(U_i^1) \sim \mathcal{N}(-\frac{\sigma_{U,1}^2}{2}, \sigma_{U,1}^2)$ and $\log(U_i^2) \sim \mathcal{N}(-\frac{\sigma_{U,2}^2}{2}, \sigma_{U,2}^2)$, implying $E(U_i^1|Z_i^1(t)) = 1$ and $E(U_i^2|Z_i^2(t)) = 1$ for all t occurring in the first and in the second exposure period, respectively. When accounting for the shared components of Berkson error due to individual job conditions and worker practices, the vector of variance parameters of $\log(U_i^q)$ therefore becomes $\boldsymbol{\sigma}_U^2 = (\sigma_{U,1}^2, \sigma_{U,2}^2)$. In a first attempt to account for shared Berkson error in the French cohort of uranium miners, we assumed these variance parameters to be known and equal to the values derived by Allodji et al. (2012) [33, 34], i.e. $\sigma_{U,1}^2 = 0.93^2 = 0.86$ and $\sigma_{U,2}^2 = 0.39^2 = 0.15$. The directed acyclic graph for the full model, which results from the combination of the disease model and the measurement model $\mathcal{M}_2$ is shown in Figure 5.5.

5.1.3 The exposure model

As described in section 3.2 and in section 4.4, we need to specify an exposure model if we choose a structural approach for classical measurement error correction. The exposure model specifies a family of distributions for the true and unknown values of exposure $X_i^q(t)$. In the case of Berkson error, on the other hand, the distribution of true exposure $X_i^q(t)$ is fully specified by the measurement model. Occupational exposures [14] and in particular occupational exposures that are airborne [152] are often assumed to follow a log-normal distribution. In line with this assumption, numerous residential and occupational studies have observed a log-normal distribution for radon exposure [78, 24, 27, 26, 40, 28, 69] and it is common to make this distributional assumption for true exposure $X_i^q(t)$ when modelling measurement error in radon exposure [151, 34]. Based on these arguments, we assumed a log-normal distribution to describe true and unknown radon exposure $X_i^q(t)$ in the last period of exposure assessment in the French cohort of uranium miners, which was characterised by individual dosimetry (1983-2007) when accounting for unshared classical measurement error in this period (see measurement model $\mathcal{M}_1$ in section 5.1.2).

However, based on changes in radiation protection and on technical improvements concerning

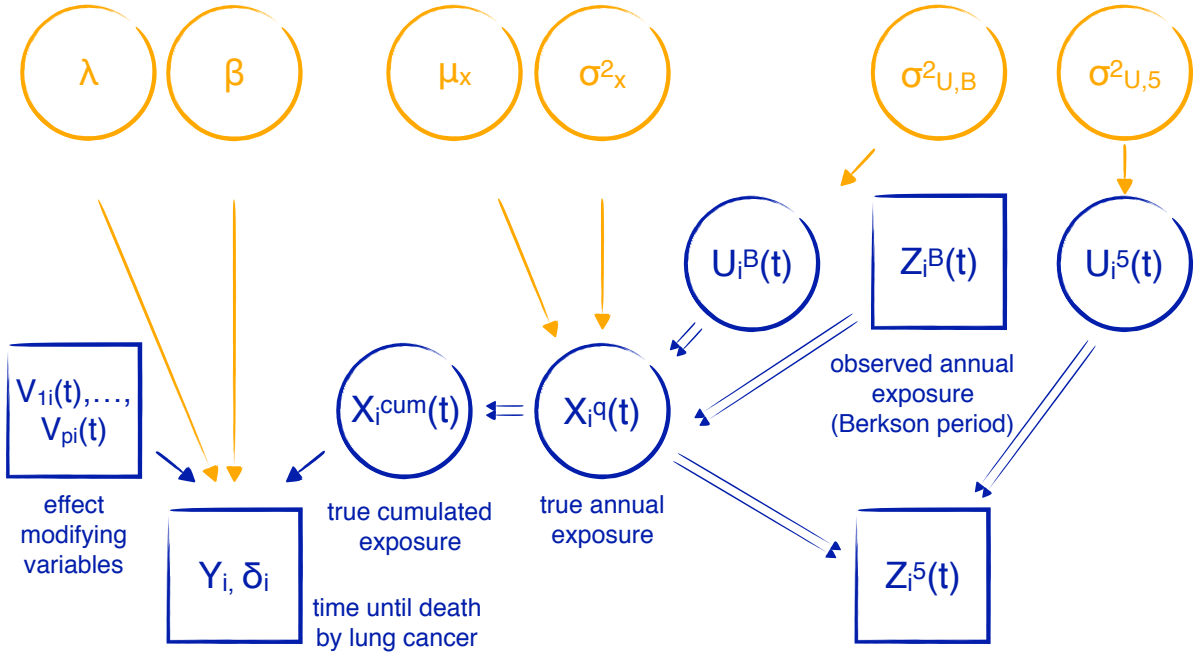

Figure 5.6: Directed acyclic graph for the full hierarchical model accounting for unshared measurement error. For the sake of clarity, we split the vector $\boldsymbol{\sigma^2_U}$ into $\boldsymbol{\sigma^2_{U,B}} = (\sigma^2_{U,1}, \sigma^2_{U,2}, \sigma^2_{U,3}, \sigma^2_{U,4})$ for the exposure periods with unshared Berkson error and $\sigma^2_{U,5}$ for the exposure period with unshared classical. Likewise, we denoted $Z_i^B(t)$ the observed exposure values for the exposure periods with unshared Berkson error and $Z_i^5(t)$ the observed exposure values for the exposure period with unshared classical error.

the work in the mines, the true radon exposure can be assumed to have decreased in the period

between 1983 and 2007. To allow different exposure distribution parameters depending on the period of exposure, we modelled the mean and variance parameters $\mu_{x,p}$ and $\sigma^2_{x,p}$ of the normal distribution for $\log(X_i^5(t))$ as a function of period of exposure p. Period p takes values in $\{1, 2, 3, 4, 5\}$ corresponding to the exposure periods 1983-1984, 1985-1986, 1987-1989, 1990-1994 and 1995-2007, respectively. These exposure periods were chosen so as to obtain a comparable number of measurements in each period. $\boldsymbol{\mu}_x = (\mu_{x,1}, \mu_{x,2}, \mu_{x,3}, \mu_{x,4}, \mu_{x,5})$ therefore denotes the vector of unknown mean parameters of the natural logarithm of true exposure and $\boldsymbol{\sigma}^2_x = (\sigma^2_{x,1}, \sigma^2_{x,2}, \sigma^2_{x,3}, \sigma^2_{x,4}, \sigma^2_{x,5})$ denotes the corresponding vector of variance parameters. The DAG for the full model, which results of the combination of the disease model, measurement model $\mathcal{M}_1$ and the exposure model (described in section 5.13), is shown in Figure 5.6.

5.1.4 The dose model

As described in section 2.3, the concept of detriment-weighted absorbed lung dose is more suitable to reflect the potential health effects of exposure to radon and its progeny than exposure values expressed in working level months (WLM). However, there are a certain number of uncertain input parameters involved in the calculation of absorbed lung dose. One of the objectives of this work was to account simultaneously for the uncertainty on these input parameters and for exposure measurement error when estimating the risk of lung cancer mortality associated with radon exposure in the French cohort of uranium miners. For this purpose, we both needed to determine a dose model which allows to calculate detriment-weighted absorbed lung dose and to specify prior distributions that reflect our uncertainty on these input parameters. For this purpose, we needed to derive a simplified but realistic mathematical expression of the Human Respiratory Tract Model (HRTM), which is commonly used to model detriment-weighted absorbed lung dose as presented in section 2.3.2. Our goal was to be able to both easily combine this so-called "dose model" with the disease model, the measurement model and the exposure model as defined previously and to be able to conduct Bayesian inference for the resulting full hierarchical model. To this end, we additionally had to specify prior distributions for the input parameters intervening in this dose model. To specify the dose model, as well as the prior distributions on the uncertain input parameters intervening in this model, were worked in collaboration with experts on the calculation of radiation doses for radon progeny both at the laboratory of internal dosimetry at IRSN and at Public Health England (PHE) [405, 37, 38, 116, 39, 59, 60, 104]. A general defintion of the approximated detriment-weighted absorbed dose $D_i(t)$ (in mGy) received by miner i at time t can be given by:

$$D_i(t) = \frac{br_i(t)}{cbr} \cdot (fp_i(t) \cdot cunat_i(t) + (1 - fp_i(t)) \cdot cat_i(t)) \cdot X_i(t) \tag{5.15}$$

where cbr is a proportionality constant that is equal to 1.2 $\mathrm{m^3 h^{-1}}$, $X_i(t)$ is the true radon (progeny) exposure that miner i received at time t, $br_i(t)$ denotes the average breathing rate of miner i at time t in $\mathrm{m^3 h^{-1}}$ and $fp_i(t)$ is the unattached fraction of radon progeny to which miner i was exposed at time t (see section 2.3.3 for a more detailed presentation of these input parameters). The two dose conversion coefficients $cunat_i(t)$ and $cat_i(t)$ (expressed in mGy/WLM) are a function of the Activity Median Diameter (AMD) of the unattached ($Auat_i(t)$) and the attached radon progeny ($Aat_i(t)$) to which miner i was exposed to at time t, respectively. Public Health England provided us with a certain number of AMD values given in nanometer (nm)

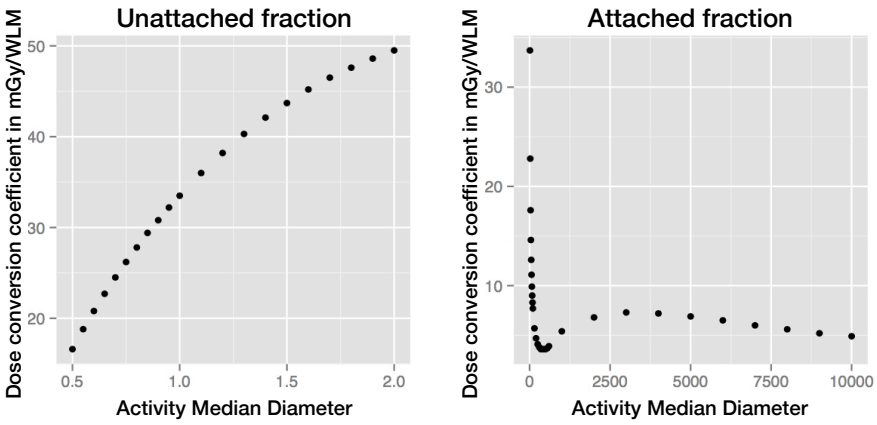

Figure 5.7: Values of the Activity Median Diameter (AMD) and the corresponding dose conversion coefficients of unattached and attached radon progeny.

for the unattached radon progeny $Auat$ and for the attached radon progeny Aat as well as with the corresponding dose conversion coefficients, which were obtained through a more complex and more realistic implementation of the HRTM. Figure 5.7 shows these values for the attached and unattached fraction of radon progeny. We used cubic splines to interpolate the relation between these quantities, resulting in the interpolated functions $\hat{g}_1$ and $\hat{g}_2$ plotted in Figure 5.8. These cubic spline functions were fitted in Python and subsequently used to determine a dose conversion coefficient corresponding to a given value of the AMD for attached and unattached radon progeny, respectively.

In order to achieve a more parsimonious version of the dose model defined in (5.15), we made

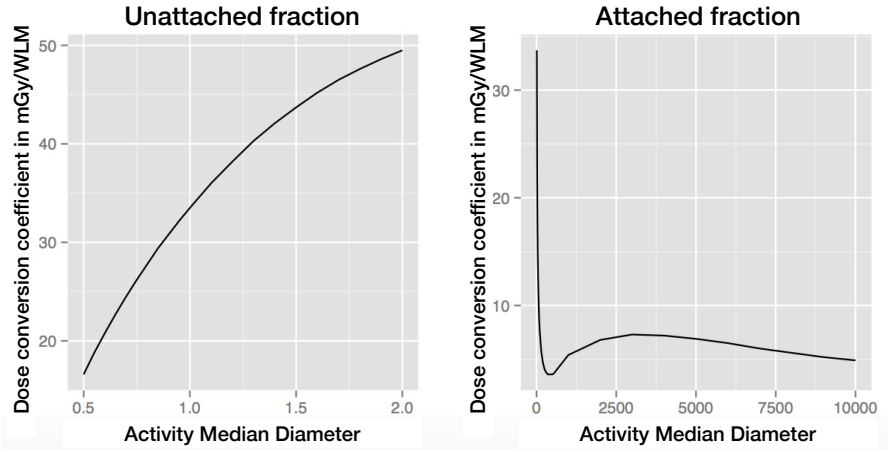

Figure 5.8: Cubic spline interpolation of the relation g_1 and g_2 between the Activity Median Diameter (AMD) and the corresponding dose conversion coefficients of unattached and attached radon progeny.

the hypothesis that the uncertain input parameters in the dose model, i.e. $br_i(t)$, $fp_i(t)$, $Aat_i(t)$ and $Auat_i(t)$ remain constant for all exposure years characterised by a manual type of work (i.e. until 1977) and all exposure years characterised by mechanised type of work (i.e. after

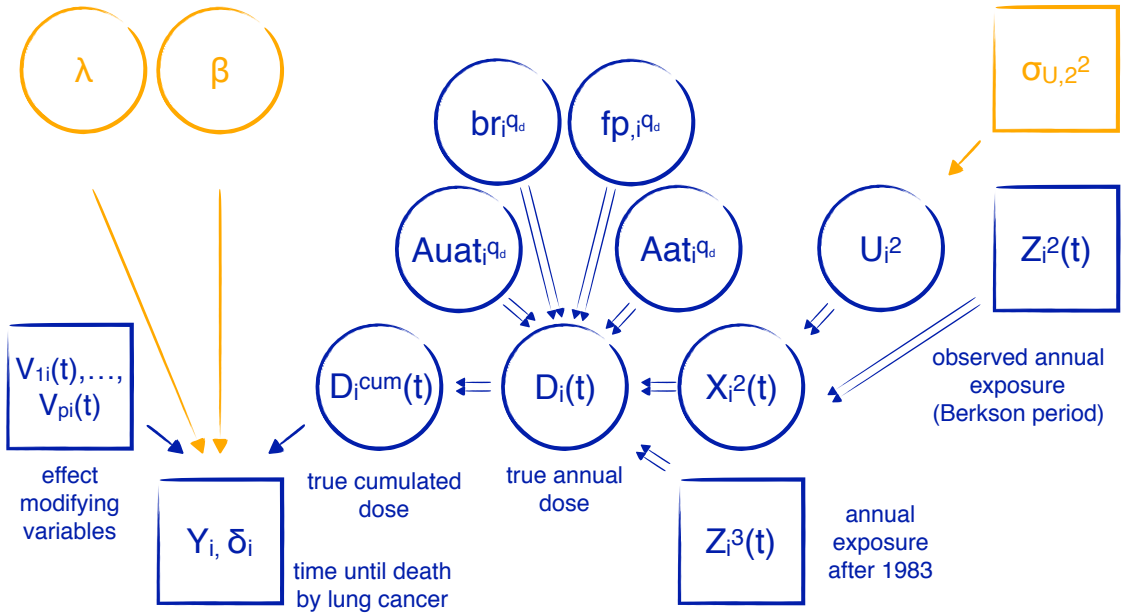

Figure 5.9: DAG for the full hierarchical model accounting for uncertainty on the input parameters in the dose model and for shared Berkson error for the period 1956-1982 while neglecting unshared classical measurement error for the period 1983-2007 in the post-55 cohort.

1977)[1]. Indeed, the working conditions of a miner in the French cohort of uranium miners can be supposed to have stayed more or less constant over the years in each of these two periods. In the French cohort of uranium miners, only the working conditions in the years after 1955 have been characterised thoroughly with the aim to inform dosimetric calculations [116, 88]. When accounting for dose uncertainty, we therefore restricted our analyses to the so-called "post-55" cohort, i.e. all miners in the French cohort of uranium miners employed after 1955 (see section 2.2.2). With this restriction and under the hypothesis that the input parameters for a miner i remain constant in a dose period q_d, where $q_d = 1$ for exposures received until 1977 and $q_d = 2$ for exposures received after 1977, the approximated detriment-weighted absorbed lung dose $D_i^{q_d}(t)$ received by miner i at time t corresponding to dose period q_d can be given by the following model, which we will denote as dose model $A1$:

$$D_i(t) = \frac{br_i^{q_d}}{cbr} \cdot (fp_i^{q_d} \cdot cunat_i^{q_d} + (1 - fp_i^{q_d}) \cdot cat_i^{q_d}) \cdot X_i^q(t) \tag{5.16}$$

where $q \in \{2, 3\}$ corresponding to the exposure years between 1956 and 1982 and between 1983 and 2007, respectively. In order to account simultaneously for the uncertainty on the input parameters intervening in dose calculation and for shared exposure measurement error, we combined dose model $A1$ with the following measurement model:

$$\boldsymbol{\mathcal{M}_2^*}: \begin{cases} X_i^2(t) = Z_i^2(t) \cdot U_i^2 \\ Z_i^3(t) = X_i^3(t). \end{cases}$$

that accounts for Berkson error shared within miners in the second exposure period, that was characterised by a prospective method of group-exposure assessment via ambient measurements.

[1]See Allodji et al. (2012) [33] for more details on the changes in working conditions in French uranium mines.

The DAG of the full hierarchical model, which results of the combination of the disease model, the measurement model $\mathcal{M}_2^*$ and the dose model is shown in Figure 5.9. It is interesting to note that, in contrast to the disease model, the measurement model and the exposure model, which have been presented earlier, the dose model we considered in this work is a deterministic model. In other words, a given set of input parameters will always result in the same dose value $D_i^{q_d}(t)$. The epistemic uncertainty on the input parameters for each miner i, on the other hand, is expressed via prior distributions on these unknown quantities that we will define in the next section. In Figure 5.9 the deterministic dependence between the dose estimate $D_i^{q_d}(t)$ and the input parameters is expressed via double arrows.

5.1.5 Defining prior distributions

As described in section 4.1 and illustrated in the different DAGS presented in the last section (see Figure 5.6, Figure 5.5 and Figure 5.9), we need to specify prior distributions in order to conduct Bayesian inference for all unknown quantities. Table 5.1 summarises the parameters in all sub-models.

Table 5.1: Summary of the unknown quantities in the sub-models for which prior distributions have to be specified

Sub-model	Unknown quantity	Description	Details
Disease model			
	$\boldsymbol{\beta} = (\beta_r)_{r=1,\dots,R}$	Risk association	$R = 1$ for the simple linear models ($\mathcal{D}_1$, $\mathcal{D}_2$) $R = 2$ for the piecewise linear models ($\mathcal{D}_5$, $\mathcal{D}_6$) and the period of exposure model ($\mathcal{D}_3$) $R = 3$ for the time since exposure model ($\mathcal{D}_4$)
	$\boldsymbol{\lambda} = (\lambda_j)_{j=1,\dots,J}$	Baseline hazard	$J = 4$ age periods: < 40, 40-55, 55-75 and 75-85 years
Measurement model			
	$\boldsymbol{\sigma_U^2} = (\sigma_{U,q}^2)_{q=1,\dots,Q}$	Measurement error variance	$Q = 5$ for unshared measurement error ($\mathcal{M}_1$) with calendar periods (46-55, 56-74, 75-77, 78-82, 83-07) $Q = 3$ for shared measurement error ($\mathcal{M}_2$) with calendar periods (46-55, 56-82, 83-07)
Exposure model			
(unshared classical error for 1983-2007)	$\boldsymbol{\mu_x} = (\mu_{x,p})_{p=1,\dots,P}$	Mean of the logarithm of true exposure	$P = 5$ calendar periods (83-84, 85-86, 87-89, 90-94, 95-07)
	$\boldsymbol{\sigma_x^2} = (\sigma_{x,p}^2)_{p=1,\dots,P}$	Variance of the logarithm of true exposure	
Dose model			
(for exposure values received after 1956)	$\boldsymbol{fp} = (fp_i^{q_d})_{i=1,\dots,3377}^{q_d=1,\dots Q_d}$	Unattached fraction of radon progeny	
	$\boldsymbol{Aat} = (Aat_i^{q_d})_{i=1,\dots,3377}^{q_d=1,\dots Q_d}$	Activity Median Diameter of attached radon progeny	$Q_d = 2$ calendar periods (56-82, 83-07)
	$\boldsymbol{Auat} = (Auat_i^{q_d})_{i=1,\dots,3377}^{q_d=1,\dots Q_d}$	Activity Median Diameter of unattached radon progeny	
	$\boldsymbol{br} = (br_i^{q_d})_{i=1,\dots,3377}^{q_d=1,\dots Q_d}$	Average breathing rate	

Disease model

We chose independent prior distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$. In particular, we opted for vague prior distributions for the vector of unknown risk coefficients $\boldsymbol{\beta}$ in the form of centred normal distributions with large variances (10^4). As mentioned in section 5.1.1 and following Ibrahim et al. (2001) [380], we used the external data on the yearly lung cancer mortality rate in French males between 1968 and 2005 [400] to specify independent and informative gamma priors for the components of $\boldsymbol{\lambda}$. In accordance with these mortality rates, we defined gamma priors $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \beta_{0j})$ with α_{0j} taking values 23.66, 35.53, 88.10 and 29.75 and β_{0j} taking values $4.90 \cdot 10^8, 2.58 \cdot 10^7, 1.61 \cdot 10^7$ and $3.25 \cdot 10^6$ for the four time intervals $j = 1, 2, 3, 4$.
In order to test the influence of these informative prior distributions, we estimated the parameters of the disease model assuming flat uniform distributions between zero and one for the parameters λ_2, λ_3 and λ_4. For the parameter λ_1 describing baseline hazard before 40 years, we always assumed the informative gamma distribution defined earlier, since only one miner died of lung cancer before 40 years in the cohort. As there was very little information on this parameter in the data, it would have been inadequate to estimate its value based on a flat uniform prior distribution.

Measurement model

Concerning the parameters $\sigma^2_{U,1}, \sigma^2_{U,2}, \sigma^2_{U,3}, \sigma^2_{U,4}$ and $\sigma^2_{U,5}$, describing the variance of the log-transformed values of measurement error in the five exposure periods, we did a parameter transformation. Instead of modelling these variance parameters directly, we specified informative prior distributions on their square roots, i.e. on the corresponding standard deviation parameters $\boldsymbol{\sigma}_U$. We assigned independent normal distributions that were centred at the values 0.93, 0.47, 0.42, 0.33 and 0.10, respectively. These values correspond to the values determined by Allodji et al. (2012) [33, 34] when characterising exposure measurement error in the French cohort of uranium miners. The standard deviation parameters of the prior distributions on $\boldsymbol{\sigma}_U$were set to $0.03, 0.005, 0.005, 0.005$ and 0.0005.

Exposure model

To account for unshared classical measurement error after 1983 through measurement model $\mathcal{M}_1$, we have to specify prior distributions on the vector of values describing the mean $\boldsymbol{\mu}_x$ and the variance $\boldsymbol{\sigma}^2_x$ of the log-transformed true exposure values $X^5_i(t)$ received in this period. These prior distributions were defined in accordance with data on radon exposure in a sub-cohort of the Wismut cohort consisting of 11.000 miners that were first employed by the Wismut company between 1971 and 1989 [406]. The miners in this sub-cohort can be supposed to have been exposed to radon in a similar way as the miners in the French cohort after 1983, since their exposure conditions were subject to similar technological developments and the same guidelines for radiation protection. Among exposed miners in this sub-cohort, the median value of cumulated radon exposure was around 5 WLM and mean radon exposure was 9 WLM [406]. If we divide these values by the average duration of exposure in this cohort, which was seven

years, we obtain a yearly median exposure of 0.71 WLM and a yearly mean exposure of 1.29 WLM. Using the properties of the log-normal distribution we can deduce a geometric mean of -0.34 and a geometric variance of 1.18 for this sub-cohort.
We therefore assumed the prior expected value of $\boldsymbol{\sigma}_x^2$ in the French cohort of uranium miners to be 1.18 for all elements of $\boldsymbol{\sigma}_x^2$. To respect this condition, as well as a 95th percentile of 2.89 (which was determined via plots of the resulting log-normal distributions for different values of $\boldsymbol{\mu}_x$) we chose an inverse gamma prior for $\boldsymbol{\sigma}_x^2$ with shape parameter $\alpha_{\sigma_x^2} = 1.75$ and scale parameter $\beta_{\sigma_x^2} = 0.88$ for all elements of $\boldsymbol{\sigma}_x^2$. The resulting prior credible intervals for $\boldsymbol{\sigma}_x^2$ are $CI_{95\%}$: [0.17; 5.22]. Inverse gamma distributions as prior distributions for variance parameters are a classical choice for Bayesian inference to respect their positivity and to obtain analytically tractable full conditional distributions for these parameters.
The steady decrease in annual mean radon exposure that is observed in cohorts of underground miners after 1965 make the elicitation of the prior distribution for $\boldsymbol{\mu}_x$ based on the sub-cohort of the German cohort of uranium miners more difficult. Indeed, all information on radon exposure based on this cohort is valid for the period between 1971 and 1989. The exposure model we want to specify, on the other hand, concerns the period between 1983 and 2007 in the French cohort, characterised by classical measurement error. To account for this, we did not directly centre the prior distribution for $\boldsymbol{\mu}_x$ around its corresponding value observed in the Wismut sub-cohort. This prior is only required to have a non-negligible prior probability for the median radon exposure value to be greater or equal to the median value observed for the Wismut sub-cohort. At the same time we required that there should be a non-negligible probability for median radon exposure in the French cohort after 1983 to be more than 10 times smaller than in the German cohort to account for the possibility of a gradual decrease in yearly radon exposure in uranium mines. In order to respect these two conditions, we chose a normal prior distribution for all components of $\boldsymbol{\mu}_x$ with mean -1.44 and variance 10.24, corresponding to a 95% credible interval ($\text{CI}_{95\%}$) [-7.71; 4.83]. The influence of the chosen priors on parameter estimation was assessed by comparing the obtained results with results obtained when assuming more informative prior distributions for all components of $\boldsymbol{\mu}_x$ via a normal distribution with mean -1.44 and a variance of 1.44 ($\text{CI}_{95\%}$: [-3.79; 0.91]) and an inverse gamma distribution with $\alpha_{\sigma_x^2} = 3.11$ and $\beta_{\sigma_x^2} = 2.47$ ($\text{CI}_{95\%}$: [0.34; 3.70]).

Dose model

Concerning the specification of information prior distributions on the uncertain input parameters in the dose model, we worked in collaboration with dosimetrists at IRSN and at Public Health England (PHE), as mentioned in section 5.1.4. In the context of this collaboration, we specified informative prior distributions for fp^{q_d}, Aat^{q_d} and $Auat^{q_d}$ which were mainly based on measurements in mines performed by Butterweck et al. (1992) [115], Porstendörfer et al. (1999) [407], Bouland et al. (1992) [408], Bigu (1990) [409] and on the values proposed by Birchall et al. (1994) [2]. For the exposure years after the mechanisation of work in the mines, i.e. for $q_d = 2$, it was assumed that the diesel aerosol dominated the mine aerosol, resulting in a very low unattached fraction and an activity median diameter (AMD) around 200 nm for attached radon progeny [115]. This information was translated into log-normal distributions for both fp_i^2 and Aat_i^2 with a geometric mean of 0.0006 and a geometric standard deviation of 2.0 for fp_i^2 and

a geometric mean of 200 nm and a geometric standard deviation of 1.3 for Aat_i^2. Concerning the working conditions before the mechanisation (i.e. until 1977), which were characterised by an absence of diesel engines and by medium to good ventilation, it was assumed that the mine aerosol was similar to the outdoor aerosol having a higher unattached fraction and a larger AMD for the attached radon progeny [115, 410, 116]. In accordance, we assumed log-normal distributions for the unattached fraction fp_i^1 and the activity median diameter of the attached radon progeny Aat_i^1 for the first dose period with a geometric mean of 0.01 and a geometric standard standard deviation of 1.9 for fp_i^1 and a geometric mean of 250 nm and a geometric standard deviation of 1.2 for Aat_i^1. Concerning the activity median diameter of the unattached radon progeny, we assumed a uniform distribution between 0.5 nm and 1.5 nm for both dose periods, i.e. $Auat_i^{q_d} \sim \mathcal{U}[0.5 \text{ nm}; 1.5 \text{ nm}]$ $q_d \in \{1, 2\}$.
Finally, for average breathing rate of miner i in the dose period q_d, $br_i^{q_d}$, we performed a sensitivity analysis on the prior choice by assuming two alternative prior distributions. We either assumed a log-normal distribution with a geometric median 1.2 m^3h^{-1} and a geometric standard deviation of 1.3 for both dose periods, which was derived from information in the literature [411] or prior distributions based on interviews that we performed with experts on the working conditions in the French uranium mines. The methods of expert prior elicitation that we used for this purpose will be described in the next section.

5.2 The elicitation of prior information for average breathing rate by expert knowledge

Contrary to the unknown parameters in the disease and in the exposure model, the uncertain input parameters intervening in the dose model (A1) i.e. $br_i^{q_d}$, $fp_i^{q_d}$, $Aat_i^{q_d}$ and $Auat_i^{q_d}$ (with $q_d \in \{1, 2\}$ as defined in section 5.1.4) are likely to be only poorly informed by the observed data. It therefore appears to be indispensable to specify informative prior distributions for the uncertain input parameters intervening in the dose model. As described in section 5.1.5., we used information available in the literature to derive informative prior distributions for this purpose. However, it can be argued that special attention should be paid to the definition of prior distributions concerning average breathing rate. First of all, this input parameter is likely to be the most influential quantity in the dose model (apart from radon exposure $X_i^q(t)$). More importantly, the prior information available in the literature on this unknown input parameter is mainly based on measurements that were performed for workers in metal mines in Tajikistan and gold mines in South Africa [2, 412, 102, 411]. It is questionable to what extent the working conditions in these mines can be extrapolated to the working conditions in French uranium mines. Finally, the average breathing rate of a miner can be modelled as a function of observable quantities, contrary to the other uncertain input parameters intervening in the dose model. Consequently, we decided to derive informative prior distributions on the average breathing rate of miner a miner based on the elicitation of expert knowledge on the working conditions in French uranium mines.
In section 5.1.4, we assumed $br_i(t) = br_i^{q_d}$ with $q_d = \{1, 2\}$. When deriving a prior distribution on average breathing rate informed by the knowledge of experts on French uranium mines, we assumed a more parsimonious model. We assumed that $br_i(t) = br^l$ for all miners i and all

exposure years t corresponding to a working condition $l \in \{1, 2, 3, 4, 5, 6\}$. Each working condition was defined by a dose period (until or after 1977) and a type of work (hewer, underground miner and open pit miner). Thereby, we assume that the average breathing rate is the same for all miners in this working condition as they can be supposed to perform the same tasks. To define the average breathing rate of a French uranium miner in working condition l, i.e. br^l, it is common to multiply the proportion of time he spent in a certain level of physical activity a in working condition l denoted $P_{a,l}$ by a value $\bar{br}_a$ corresponding to the average breathing rate for this activity, to finally sum over all activities a, where $a \in \{1, 2, 3\}$ corresponding to sitting, light exercise and heavy exercise:

$$br^l = \bar{br}_1 \cdot P_{1,l} + \bar{br}_2 \cdot P_{2,l} + \bar{br}_3 \cdot P_{3,l} \tag{5.17}$$

with $\sum_{a=1}^{3} P_{a,l} = 1$ and $l \in \{1, 2, 3, 4, 5, 6\}$. This approach, which is also known as the time-activity-ventilation approach [411], has the advantage that the proportion of time a worker in a given working condition commonly spent in a certain level of physical activity is a variable that can be derived by considering the different tasks performed in a specific mine. The average breathing rate parameters $\bar{br}_a$, corresponding to the different levels of physical activity, on the other hand, can safely be extrapolated from one worker population to the other. Finally, the definition of average breathing rate via the time-activity-ventilation approach relates it to a quantity with three components $P_{a,l}$ that are observable by experts and in line with the general recommendations on prior elicitation presented in section 4.3.3 according to which we should seek to elicit expert information only on observable quantities. In order to make the quantities that were to be evaluated by the experts even more palpable to them, we decided to ask for the time a miner in working condition l spent in the different levels of physical activity a during a working day, instead of asking for a proportion. We specified in the beginning of the elicitation exercise that, in our definition, a working day consisted of eight hours.

In line with these arguments, we designed an elicitation task to elicit the knowledge of expert e concerning the time a miner in a working condition l spent in a given physical activity a, $S_{e,a,l}$ with $S_{e,1,l}$ corresponding to the time spent sitting $S_{e,2,l}$ to the time spent in light exercise and $S_{e,3,l}$ to the time spent in heavy exercise.

The first step to derive informative prior distributions on the average breathing rate in French uranium mines based on expert knowledge, was to identify experts, which were capable of providing us with information on the time French uranium miners spent in different levels of physical activity. To that purpose, we contacted a historian, who was familiar with the working conditions in French uranium mines. He recommended us two additional experts, who had formerly worked in a French uranium mine. In order to guarantee their anonymity, we will call these three experts A, B and C in the following. Experts A and B were employed as electrician and mechanic in a French uranium mine, respectively, and expert C was the historian familiar with the working conditions in French uranium mines.

Unfortunately, the aggregation of the opinions of multiple experts poses additional difficulties in the elicitation of expert opinion. The most intuitive solution to this problem is arguably to organise a group elicitation session during which a unique prior distribution reflecting the experience of the whole group is elicited [279]. However, this approach, which is also referred to as the behavioural approach to the aggregation of expert opinion [350], has a number of potential drawbacks. First of all, it is not always possible to organise an interview session with

all subject-matter experts at the same time. Moreover, this strategy may promote the use of heuristics and lead to overconfident judgments [350] as well as phenomena like group polarisation according to which a group may adopt more extreme positions than would individual members of the same group [346]. Finally, a group elicitation would have been inconvenient in the case of the three experts we had identified, because they were employed in the same company and expert C was higher in the working hierarchy than expert A and expert B. In the context of a group elicitation session, it is therefore questionable whether expert A and expert B would have been free to express their opinion in cases where they did not agree with expert C. Based on these arguments, we organised individual sessions of about two to three hours for the elicitation of prior information for each expert. These individual sessions also allowed us to make sure that each expert had correctly understood the elicitation task, which will be presented in section 5.2.1.

Given the data elicited by each expert e, $e \in \{1, 2, 3\}$, we derived a joint prior probability distribution $f_{e,l}$ on the three-dimensional vector $(P_{e,1,l}, P_{e,2,l}, P_{e,3,l})$ reflecting the expert's knowledge concerning the proportion of time spent sitting, in light exercise and in heavy exercise, denoted $P_{e,1,l}, P_{e,2,l}$ and $P_{e,3,l}$, respectively, for a miner in working condition l, $l \in \{1, 2, 3, 4, 5, 6\}$ with $\sum_{a=1}^{a} P_{e,a,l} = 1$. Once we had specified an individual joint prior distribution for each expert, we compared three mathematical approaches [279, 350] for the aggregation of expert opinion, which we will present in section 5.2.3. Through these aggregation methods, we obtained a unique joint prior distribution f_l on the three-dimensional vector $(P_{1,l}, P_{2,l}, P_{3,l})$ of the proportion of time a miner in working condition l spent in the different levels of physical activity a, i.e. $P_{1,l}$, $P_{2,l}$ and $P_{3,l}$ corresponding to the proportion of time spent sitting, in light exercise and in heavy exercise, respectively with $\sum_{a=1}^{a} P_{a,l} = 1$. Finally, based on f_l, we derived a unique prior distribution on the average breathing rate br^l of a miner in working condition l using the time-activity-approach (see equation 5.15). We assumed the values $\bar{br}_1 = 0.54$ m^{3}h^{-1}, $\bar{br}_2 = 1.5$ m^{3}h^{-1} and $\bar{br}_3 = 3.0$ m^{3}h^{-1} for the average breathing rate in the proportion of time spent sitting, in light exercise and in heavy exercise, respectively, as recommended by the International Commission on Radiological Protection (ICRP) [102].

5.2.1 Designing the elicitation task

As described above, the quantity of interest for which we sought to gather information through the knowledge of expert e in the elicitation task was the time $S_{e,a,l}$ a miner in working condition l ($l \in \{1, \ldots, 6\}$) spent in a given level of physical activity a ($a \in \{1, 2, 3\}$), during a workday of eight hours. In the design of the elicitation exercise, we relied on the idea of Frank Ramsey and Bruno de Finetti according to which the degree of belief of a person can be observed by their willingness to bet on their beliefs [280] (see section 4.1.1). In line with this idea, we chose an indirect elicitation task in which the expert did not explicitly have to state any quantity describing the probability distribution that most adequately reflected his uncertainty on $S_{e,a,l}$. Rather, we asked him to make a series of binary choices to indicate his preference between the spinning of a fortune wheel and a bet on $S_{e,a,l}$. In order to ensure a certain level of objectivity and reproducibility of the elicitation exercise, we developed an elicitation software. This software was coded in Python version 2.7. The use of a computerised implementation of the elicitation exercise also allowed us to give instantaneous graphical feedback on the elicited information, as

we will describe below.

Given a level of physical activity a (i.e., sitting, light exercise or heavy exercise), the expert e

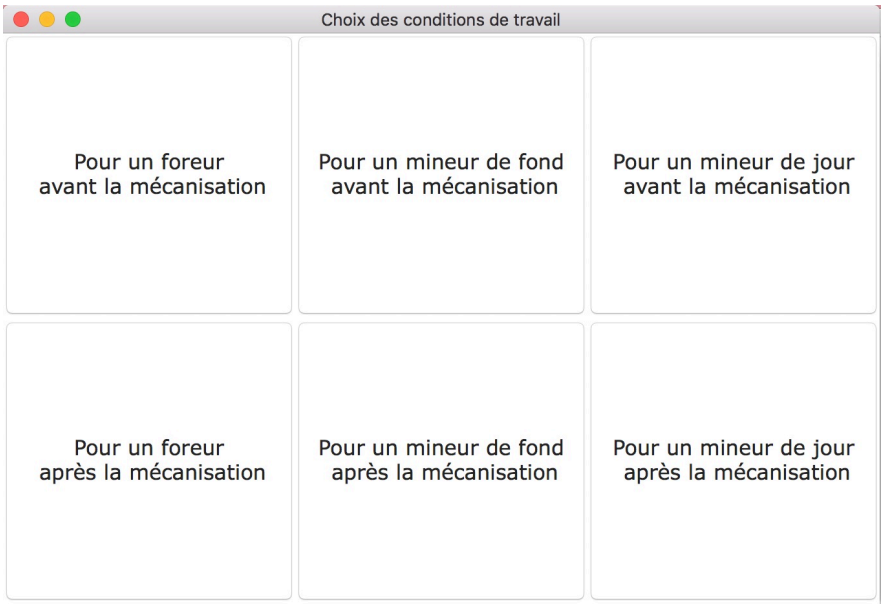

Figure 5.10: Initial screen (in French) of the developed elicitation software to choose a working condition.

was first asked to choose a working condition l via the screen shown in Figure 5.10. Once he had made this initial choice, we asked the expert e to make a sequence of binary choices in order to derive the median value $q^{0.5}_{e,a,l}$ of the time $S_{e,a,l}$ that a miner in the working condition l spent in the level of physical activity a. In the context of the indirect specification of this median value, the expert was instructed that, in order to win a hypothetical lottery of 20 euros, they would either have to turn a fortune wheel which presented an equal number of blue and beige-coloured segments or to take a bet on the time a miner in working condition l spent in the level of physical activity a. If they chose the fortune wheel, they would win the hypothetical lottery if the wheel landed on blue. If they chose the bet, they would win if the statement they were betting on was correct. Thereby, the elicitation task provides a way to measure their epistemic uncertainty (on the variable of interest $S_{e,a,l}$) via a well-calibrated and easily understandable measure of aleatory uncertainty (in the form of the fortune wheel).

Figure 5.11 shows an example screen in which experts were asked to either take the bet that an underground miner after the mechanisation spent less than one hour and a half sitting (in a working day of eight hours) or to spin the fortune wheel on the left (during a training phase that we will describe in the end of this section, we had clarified that they would have a 50% chance of winning when choosing this fortune wheel). There is evidence that a frequency interpretation of the uncertainty on a variable facilitates probability judgements for an expert (see section 4.3.3). In order to avoid cognitive biases and heuristics, we therefore supplemented the information given on the screen with a frequency interpretation of the bet on the right-hand side in the sense that we asked the experts to imagine 100 uranium miners in a given working condition and to decide whether more or less than 50 of these 100 miners spent less than one hour and a half sitting in a working day of eight hours. If they chose to spin the fortune wheel on the left, the software would propose a new value for $S_{e,a,l}$, which would be superior to the currently proposed

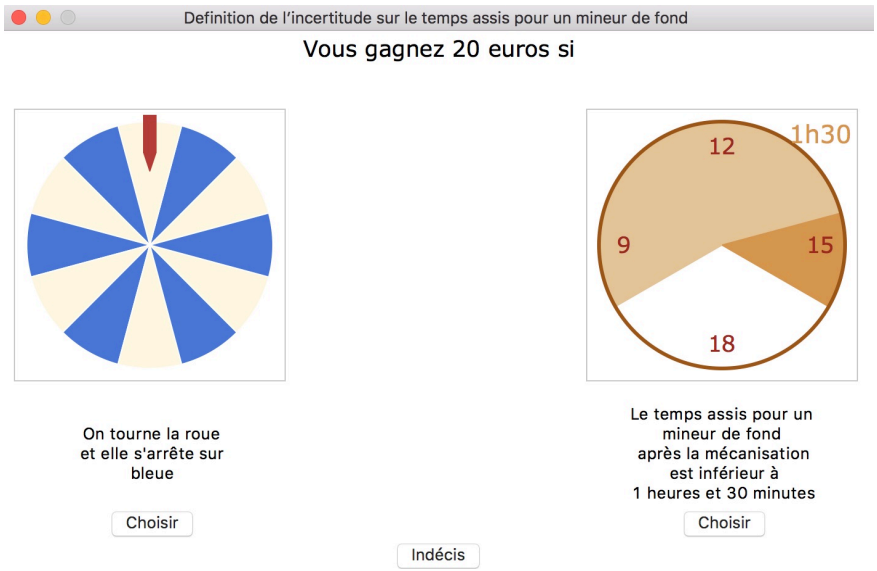

Figure 5.11: Example screen of the developed elicitation software to elicit the median value of the quantity of interest in a given working condition

value. If they chose to take the bet on right, on the other hand, the software would propose a new value for $S_{e,a,l}$, which would be inferior to the currently proposed value. In order to avoid systematic biases due to anchoring effects (see section 4.3.3), the values for $S_{e,a,l}$ were chosen randomly among the possible values every time a new value for this variable was proposed. The indirect specification of the median value $q^{0.5}_{e,a,l}$ of $S_{e,a,l}$ would be completed by one of the following two conditions:

1. The expert could choose the button "indécis" (i.e. "indifferent") for a proposed value of $S_{e,a,l}$. In this case we set $q^{0.5}_{e,a,l}$ equal to this value.

2. The expert could choose to spin the fortune wheel for a proposed value $S_{e,a,l} = v$ although he had indicated a preference for betting when the software proposed the value $S_{e,a,l} = v + 0.25$ or vice versa. In this case, we chose to interpolate the two values by setting $q^{0.5}_{e,a,l}$ equal to $v + 0.125$.

After having thus derived the median value for the time $S_{e,a,l}$ a miner in working condition l spent in a certain level of physical activity a, expert e was first presented a screen showing a fortune wheel for which only one fourth of the segments were blue as illustrated in Figure 5.12 and then a screen showing a fortune wheel for which three quarters of the segments were blue as illustrated in Figure 5.13. These screens were conceived with the aim to indirectly specify the values of the first $q^{0.25}_{e,a,l}$ and the third quartile $q^{0.75}_{e,a,l}$ of $S_{e,a,l}$, respectively. The specification of $q^{0.25}_{e,a,l}$ and $q^{0.75}_{e,a,l}$ was conducted in essentially the same manner as the specification of $q^{0.5}_{e,a,l}$. Note that the elicitation software only allowed coherent choices for the first and the third quartile in the sense that it neither proposed values for $q^{0.25}_{e,a,l}$ superior to $q^{0.5}_{e,a,l}$ nor values for $q^{0.75}_{e,a,l}$ inferior to $q^{0.5}_{e,a,l}$.

To make sure that the probability distributions that were elaborated in this way were in accordance with the subjective beliefs of the expert, it seemed very important to give an instantaneous visual feedback on the elicited quantiles. After the completion of the specification of

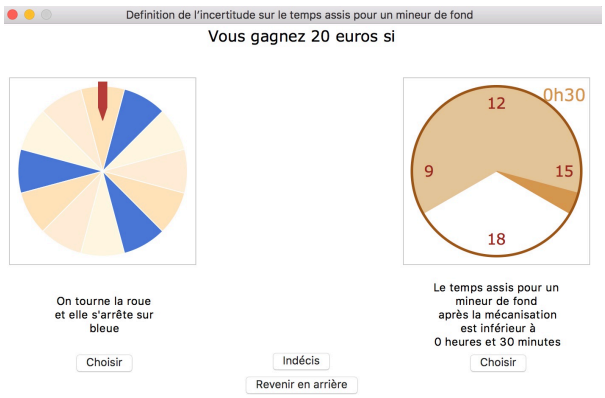

Figure 5.12: Example screen concerning the elicitation of the first quartile

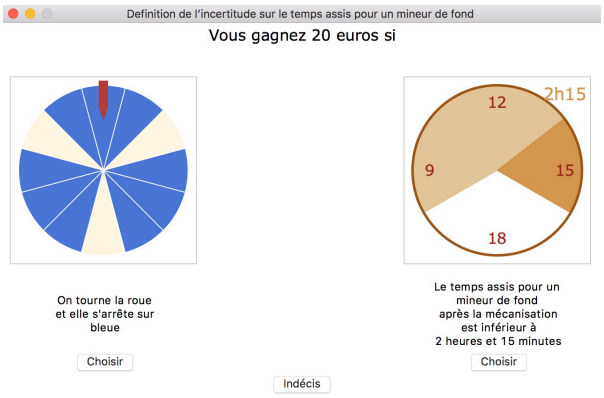

Figure 5.13: Example screen concerning the elicitation of the third quartile

$q_{e,a,l}^{0.75}$, the expert was therefore presented a screen showing two alternative probability distributions, as shown in Figure 5.14. The distribution on the left was fitted using the elicited first and second quartile whereas the distribution on the right was fitted using the elicited second and third quartile. Both probability distributions were generalised beta distributions[2] taking values between 0 and 8. The two unknown parameters of these beta distributions were estimated using a least squares minimisation method. The aim of this visual feedback was threefold. Firstly, it provided an instantaneous feedback on the elicited quantities. The probability distributions were presented as the probability distribution of $S_{e,a,l}$ for 100 miners in the working condition l. Secondly, it gave the expert the opportunity to start the specification of $q_{e,a,l}^{0.25}$, $q_{e,a,l}^{0.5}$ and $q_{e,a,l}^{0.75}$ again in the case that neither of the two distributions adequately reflected his uncertainty on $S_{e,a,l}$ by selecting the button "recommencer" (i.e. start again). Finally, the expert was asked to

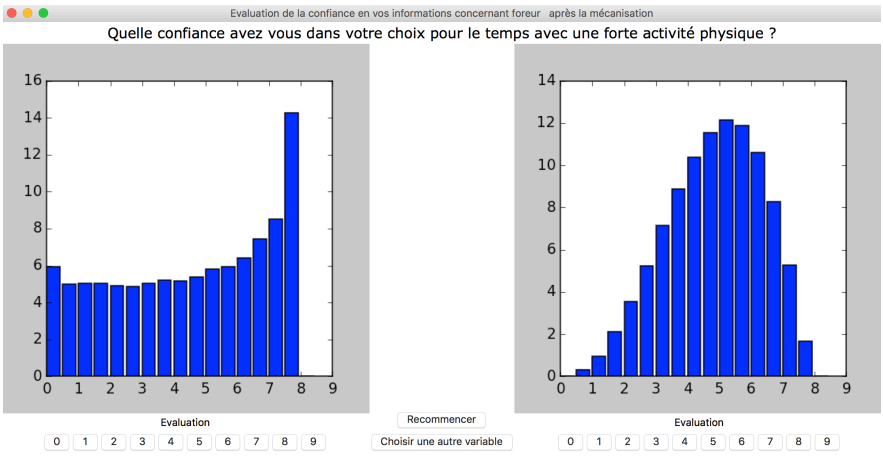

Figure 5.14: Example screen of the visual feedback provided for an expert two alternative probability distributions on $S_{e,a,l}$ derived through the elicitation task. The distribution on the left is based on the elicited first and second quartile. The distribution on the right is based on the elicited second and third quartile.

evaluate the confidence he had in the two distributions by choosing an evaluation between 0 and

[2]Let X be a random variable following a beta probability distribution with parameter a and b. Then $y_{min} + (y_{max} - y_{min})X$ follows a so called "generalised beta distribution" with parameter a and b on the interval $[y_{min}; y_{max}]$.

9 for each of the two distributions were 0 corresponded to no confidence in the plotted distribution and 9 corresponded to complete confidence in the plotted distribution. Thereby, the expert could either indicate a preference for one of the distributions or state that they equally well reflected his uncertainty on $S_{e,a,l}$. Obviously, we could have proposed a unique generalised beta distribution fitted on the three quantiles $q_{e,a,l}^{0.25}$, $q_{e,a,l}^{0.5}$ and $q_{e,a,l}^{0.75}$ in the visual feedback. However, we chose to propose two alternative distributions in this context to allow the experts to express a higher confidence in the first $q_{e,a,l}^{0.25}$ or the third quartile $q_{e,a,l}^{0.75}$ (note that the comparison of the two evaluations cannot give any information on the relative importance of the median $q_{e,a,l}^{0.5}$, as this quantity intervenes in both distributions). In the following, we will denote these evaluations taking values between 0 and 9 as $c_{e,a,l}^{1}$ for $q_{e,a,l}^{0.25}$ and $c_{e,a,l}^{2}$ for $q_{e,a,l}^{0.75}$. We used these evaluations both in the fitting of the probability distributions describing the uncertainty of each expert on the proportion of time a miner in working condition l spent in the different levels of physical activity (see section 5.2.2) and when deriving a unique prior that combined the information of all three experts (see section 5.2.3).

The training phase

In order to familiarise the expert with the format of the elicitation task, we started the prior elicitation session with a training phase. In that training phase, we adapted the example proposed by Abbas et al (2008) [1] (see section 4.3.3). We asked experts to evaluate the maximum temperature the following day. During the training session, the specification of the first, the second and the third quartile was performed as described previously. Apart from this training

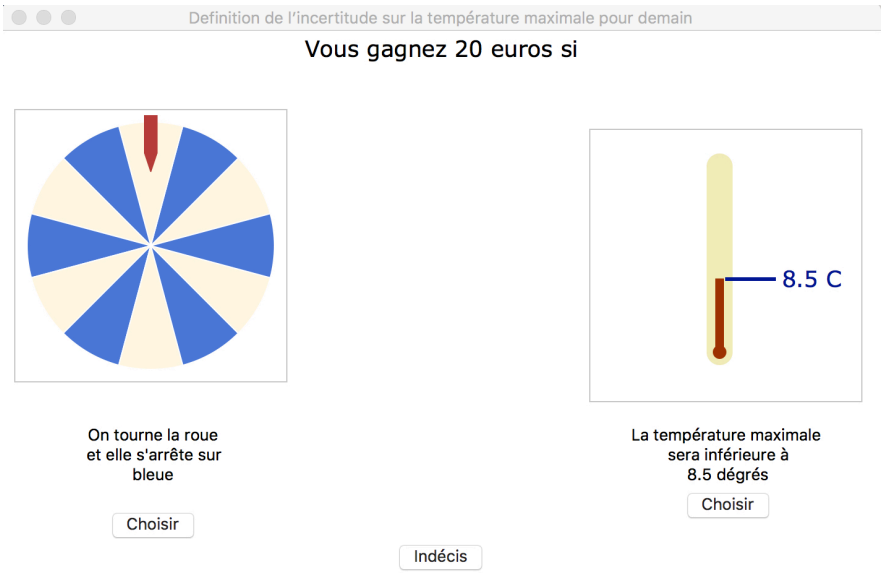

Figure 5.15: Example screen during the training phase of the elicitation task

phase concerning the maximum temperature on the following, we proposed a second training example concerning the time a person spends watching TV every day to familiarise experts with the betting on time variables. However, most experts felt sufficiently confident after the initial training phase to elicit the information on the time spent in the different levels of physical activities for a miner working in a French uranium mine.

5.2.2 Deriving a probability distribution to describe the knowledge of each expert

At the end of the elicitation process, we disposed of the elicited first, second and third quartile, $q_{e,a,l}^{0.25}$, $q_{e,a,l}^{0.5}$ and $q_{e,a,l}^{0.75}$ as well as of two evaluations $c_{e,a,l}^{1}$ and $c_{e,a,l}^{2}$ to inform us about the uncertainty expert e had on the time $S_{e,a,l}$ a miner in a working condition l spent in the physical level of activity a. For a given working condition l ($l \in \{1, \ldots, 6\}$, see Figure 5.10), our aim was to derive a unique prior distribution on the average breathing rate br^l of a miner in working condition l as a function of the joint probability distribution f_l for the three-dimensional vector $(P_{1,l}, P_{2,l}, P_{3,l})$ which corresponds to the proportion of time spent sitting, in light exercise and in heavy exercise with $\sum_{a=1}^{3} P_{a,l} = 1$ (see equation 5.17). In order to obtain f_l, an intermediate step consists in the determination of a joint probability distribution that is specific to each expert $f_{e,l}$ for the three-dimensional vector $P_{e,l} = (P_{1,e,l}, P_{2,e,l}, P_{3,e,l})$ with the constraint that $\sum_{a=1}^{3} P_{e,a,l} = 1$. At first glance, information on $P_{e,a,l}$ can easily be derived based on the information on $S_{e,a,l}$ as $P_{a,e,l} = \frac{S_{e,a,l}}{8}$. In line with this reasoning, we could simply assume that $P_{e,a,l} \sim Beta(\alpha_{e,a,l}, \gamma_{e,a,l})$ $\forall e, \forall a, \forall l$ and derive the elicited parameters $\hat{\alpha}_{e,a,l}$ and $\hat{\gamma}_{e,a,l}$ for each expert using a weighted least squares minimisation procedure based on the elicited quantiles of $S_{e,a,l}$. Indeed, we asked experts to give us information on the time spent in the different levels of physical activity for a working day of eight hours. However, this simple modelling assumption which is only based on marginal distributions would not necessarily respect the condition that the sum of the three proportions $P_{e,a,l}$ has to be equal to one. Indeed, while we elicited the knowledge of expert e on each of the variables $S_{e,1,l}$, $S_{e,2,l}$ and $S_{e,3,l}$ indicating the time spent in the different levels of physical activity independently, we know that in order to be coherent, these three quantities should sum to eight (as we defined the working day of a miner to consist of eight hours). There is some controversy on whether one should force experts to only give consistent answers in the elicitation of probability distributions [279]. As there is evidence that experts tend to overestimate each probability in a set of exhaustive and mutually exclusive scenarios, we decided to elicit the knowledge of expert e on $S_{e,a,l}$ independently for each level of physical activity a without forcing experts to respect the constraint that $\sum_{a=1}^{3} S_{e,a,l} = 8$. However, when deriving a probability distribution to reflect the knowledge of each expert e on the proportion of time $P_{e,a,l}$ spent in the different levels of physical activity a concerning a miner in working condition l, we have to respect the constraint that these proportions have to add to one. The simplest multivariate distribution respecting this condition is the Dirichlet distribution [413, 348]. The Dirichlet distribution can be seen as a generalisation of the beta distribution to more than two categories. Thus, we assumed that the three-dimensional vector $(P_{e,1,l}, P_{e,2,l}, P_{e,3,l})$ follows a Dirichlet distribution with parameters $a_{e,1,l}$, $a_{e,2,l}$ and $a_{e,3,l}$. Each of the associated marginal distributions is consequently a beta distribution with:

$$\begin{aligned} P_{e,1,l} &\sim Beta(a_{e,1,l}, a_{e,2,l} + a_{e,3,l}) \\ P_{e,2,l} &\sim Beta(a_{e,2,l}, a_{e,1,l} + a_{e,3,l}) \\ P_{e,3,l} &\sim Beta(a_{e,3,l}, a_{e,1,l} + a_{e,2,l}) \end{aligned} \tag{5.18}$$

We used these three marginal distributions to derive the elicited parameters of the proposed Dirichlet distribution for $f_{e,l}$ from the elicited quantiles.

In particular, we estimated the parameters $a_{e,1,l}$, $a_{e,2,l}$ and $a_{e,3,l}$ of the Dirichlet distribution

for expert e using a weighted least squares minimisation procedure. In other words, for each expert e and each working condition l, we looked for the set of values $(\hat{a}_{e,1,l}, \hat{a}_{e,2,l}, \hat{a}_{e,3,l})$ which minimised the sum of the weighted least square distance given by:

$$\begin{aligned}(\hat{a}_{e,1,l}, \hat{a}_{e,2,l}, \hat{a}_{e,3,l}) = \text{argmin}_{(a_1,a_2,a_3)} \sum_{a=1}^{3} & \pi^1_{e,a,l} \left(F^a_{(a_1,a_2,a_3)} \left(\frac{q^{0.25}_{e,a,l}}{8} \right) - 0.25 \right)^2 \\ & + \pi^2_{e,a,l} \left(F^a_{(a_1,a_2,a_3)} \left(\frac{q^{0.5}_{e,a,l}}{8} \right) - 0.5 \right)^2 \\ & + \pi^3_{e,a,l} \left(F^a_{(a_1,a_2,a_3)} \left(\frac{q^{0.75}_{e,a,l}}{8} \right) - 0.75 \right)^2 \end{aligned} \tag{5.19}$$

where $F^a_{(a_1,a_2,a_3)}$ denotes the cumulative distribution function of the marginal beta distribution of the a^{th} component of a Dirichlet distribution with parameter values a_1, a_2 and a_3. A grid of possible values for the three-dimensional vector (a_1, a_2, a_3) was considered numerically for this minimisation problem. This grid consisted of the values $0.05, 0.1, 0.15, \ldots, 19.9, 19.95, 20.0$, which were proposed for each of the three dimensions (these values were chosen after some initial tests that showed that it was unnecessary to proposed values greater than 20). The weights $\pi^1_{e,a,l}$, $\pi^2_{e,a,l}$ and $\pi^3_{e,a,l}$ assigned to the elicited quantiles where based on the evaluations $c^1_{e,a,l}$ and $c^2_{e,a,l}$ that the expert had provided during the elicitation task described in 5.2.1:

$$\begin{aligned} \pi^1_{e,a,l} &= \frac{\frac{2}{3} \cdot c^1_{e,a,l}}{\sum_{a=1}^{3}(c^1_{e,a,l} + c^2_{e,a,l})} \\ \pi^2_{e,a,l} &= \frac{\frac{1}{3} \cdot (c^1_{e,a,l} + c^2_{e,a,l})}{\sum_{a=1}^{3}(c^1_{e,a,l} + c^2_{e,a,l})} \\ \pi^3_{e,a,l} &= \frac{\frac{2}{3} \cdot c^2_{e,a,l}}{\sum_{a=1}^{3}(c^1_{e,a,l} + c^2_{e,a,l})} \end{aligned} \tag{5.20}$$

For a given working condition l, the elicited data consist of the three parameters of the Dirichlet distribution $f_{e,l}$ obtained for each expert, i.e. $(\hat{a}_{e,1,l}, \hat{a}_{e,2,l}, \hat{a}_{e,3,l})$ for $e \in \{1, 2, 3\}$.

5.2.3 Combining the information of several experts

For a given working condition l, our aim was to obtain a unique joint distribution f_l for $P_l = (P_{1,l}, P_{2,l}, P_{3,l})$ corresponding to the proportion of time spent sitting, in light exercise and in heavy exercise. In order to obtain this unique prior distribution, we need a method to combine the information provided by the three experts. For this purpose, we will use three different methods: Averaging, linear pooling and a Supra-Bayesian approach. Averaging and linear pooling are the most popular approaches for the combination of expert opinion [414], but they are known to understate and to overstate the variability of opinions between experts, respectively [344]. In the context of the Supra-Bayesian approach for the combination of expert opinion, the information elicited by the different experts is treated as data. In this approach, the prior distributions for breathing rate will depend on the posterior distribution on the unknown parameters in a probability model describing these elicited data.

For these three methods for the combination of expert opinion, we will use weights based on the confidence the experts indicated concerning the information they provided. The relative weights

$\pi_{1,l}, \pi_{2,l}, \pi_{3,l}$ assigned to each expert e, $e \in \{1,2,3\}$ with $\pi_{1,l} + \pi_{2,l} + \pi_{3,l} = 1 \ \forall l \in \{1,2,...6\}$ were again based on the evaluations $c^1_{e,a,l}$ and $c^2_{e,a,l}$ that the experts had provided during the elicitation exercise:

$$\begin{aligned}
\pi_{1,l} &= \frac{\sum_{a=1}^{3}(c^1_{1,a,l} + c^2_{1,a,l})}{\sum_{e=1}^{3}\sum_{a=1}^{3}(c^1_{e,a,l} + c^2_{e,a,l})} \text{ for expert 1} \\
\pi_{2,l} &= \frac{\sum_{a=1}^{3}(c^1_{2,a,l} + c^2_{2,a,l})}{\sum_{e=1}^{3}\sum_{a=1}^{3}(c^1_{e,a,l} + c^2_{e,a,l})} \text{ for expert 2} \\
\pi_{3,l} &= \frac{\sum_{a=1}^{3}(c^1_{3,a,l} + c^2_{3,a,l})}{\sum_{e=1}^{3}\sum_{a=1}^{3}(c^1_{e,a,l} + c^2_{e,a,l})} \text{ for expert 3.}
\end{aligned} \tag{5.21}$$

Note that the relative weights $\pi_{e,l}$ which reflect the confidence of each expert e for working condition l are independent of the relative weights $\pi^1_{e,a,l}, \pi^2_{e,a,l}, \pi^3_{e,a,l}$ derived for each expert and each elicited quantile and used in 5.2.2 as they use the information contained in the evaluations $c^1_{e,a,l}$ and $c^2_{e,a,l}$ differently. In the following, we will present in detail the three methods we considered and compared for the combination of expert opinion.

Averaging

Consider a working condition l with $l \in \{1,2,3,4,5,6\}$. The idea behind averaging is to assume a unique joint probability distribution f^{aver}_l for the three-dimensional vector $P_l = (P_{1,l}, P_{2,l}, P_{3,l})$, as a Dirichlet distribution for which the parameters $\hat{a}_{1,l}, \hat{a}_{2,l}, \hat{a}_{3,l}$ are simply obtained by averaging the elicited data for each of the components. In cases, where relative weights reflecting the confidence of the experts are available, it is possible to take a weighted average so that the elicited data of an expert with high confidence are more influential than the elicited data of an expert with low confidence. In our situation, we assumed the following unique prior probability distribution f^{aver}_l for $P_l = (P_{1,l}, P_{2,l}, P_{3,l})$:

$$\begin{aligned}
\hat{a}_{1,l} &= \pi_{1,l}\hat{a}_{1,1,l} + \pi_{2,l}\hat{a}_{2,1,l} + \pi_{3,l}\hat{a}_{3,1,l} \\
\hat{a}_{2,l} &= \pi_{1,l}\hat{a}_{1,2,l} + \pi_{2,l}\hat{a}_{2,2,l} + \pi_{3,l}\hat{a}_{3,2,l} \\
\hat{a}_{2,l} &= \pi_{1,l}\hat{a}_{1,3,l} + \pi_{2,l}\hat{a}_{2,3,l} + \pi_{3,l}\hat{a}_{3,3,l}
\end{aligned} \tag{5.22}$$

As pointed out by Albert et al. (2012), averaging emphasises the consensus on the elicited quantities [344].

Linear pooling

Consider a working condition l with $l \in \{1,2,3,4,5,6\}$. Contrary to averaging, linear pooling emphasises the diversity [344] in the opinions of the different experts by assuming that the unique joint prior distribution f^{pool}_l for the three-dimensional vector $P_l = (P_{1,l}, P_{2,l}, P_{3,l})$ is a mixture of three Dirichlet distributions $f_{e,l}$ for which the parameters are the elicited data provided by each expert e, $e \in \{1,2,3\}$ with

$$f^{\text{pool}}_l = \pi_{1,l} f_{1,l} + \pi_{2,l} f_{2,l} + \pi_{3,l} f_{3,l} \tag{5.23}$$

where $f_{1,l}$, $f_{2,l}$ and $f_{3,l}$ are Dirichlet distributions based on the elicited data in the following way:

$$\begin{aligned} f_{1,l} &= \text{Dirichlet}(\hat{a}_{1,1,l}, \hat{a}_{1,2,l}, \hat{a}_{1,3,l}) \\ f_{2,l} &= \text{Dirichlet}(\hat{a}_{2,1,l}, \hat{a}_{2,2,l}, \hat{a}_{2,3,l}) \\ f_{3,l} &= \text{Dirichlet}(\hat{a}_{3,1,l}, \hat{a}_{3,2,l}, \hat{a}_{3,3,l}). \end{aligned} \tag{5.24}$$

Thereby, linear pooling accumulate the plausibility of all values across experts [344].

A Supra-Bayesian approach for the combination of expert opinion

Consider a working condition l with $l \in \{1, 2, 3, 4, 5, 6\}$. Similarly to averaging, we consider in this Supra-Bayesian approach that the joint probability distribution f_l^{Supra} for the three-dimensional vector $P_l = (P_{1,l}, P_{2,l}, P_{3,l})$ is given by

$$(P_{1,l}, P_{2,l}, P_{3,l}) \sim Dirichlet(a_{1,l}, a_{2,l}, a_{3,l}). \tag{5.25}$$

The global idea of this Supra-Bayesian approach is to estimate the unknown parameters $a_{1,l}$, $a_{2,l}$ and $a_{3,l}$ in a Bayesian framework using the elicited data $(\hat{a}_{e,1,l}, \hat{a}_{e,2,l}, \hat{a}_{e,3,l})$ for $e \in \{1, 2, 3\}$ (see section 5.2.2). To this end, we need to assume a parametric model that links the elicited data $(\hat{a}_{e,1,l}, \hat{a}_{e,2,l}, \hat{a}_{e,3,l})$ to the unknown parameters $(a_{1,l}, a_{2,l}, a_{3,l})$. In order to deal with more interpretable and tractable data, we transformed the nine elicited quantities and rather considered the following elicited data in this approach:

$$\begin{aligned} \hat{\Sigma}_{e,l} &= \hat{a}_{e,1,l} + \hat{a}_{e,2,l} + \hat{a}_{e,3,l} \\ \hat{\alpha}_{e,1,l} &= \frac{\hat{a}_{e,1,l}}{\hat{\Sigma}_{e,l}} \\ \hat{\alpha}_{e,2,l} &= \frac{\hat{a}_{e,2,l}}{\hat{\Sigma}_{e,l}} \end{aligned} \tag{5.26}$$

with $\hat{\alpha}_{e,1,l}$ and $\hat{\alpha}_{e,2,l}$ taking values in $[0, 1]$ and $\hat{\Sigma}_{e,l}$ taking strictly positive real values. Note that $\hat{\alpha}_{e,a,l}$ is simply the elicited expected proportion of time spent in activity a based on the information given by expert e. Note also that $\hat{\alpha}_{e,1,l} + \hat{\alpha}_{e,2,l} + \hat{\alpha}_{e,3,l} = 1$ where $\hat{\alpha}_{e,3,l} = \frac{\hat{a}_{e,3,l}}{\hat{\Sigma}_{e,l}}$. We assume that the elicited data $(\hat{\alpha}_{e,1,l}, \hat{\alpha}_{e,2,l}, \hat{\alpha}_{e,3,l})$ follow the following marginal probability distributions:

$$(H1) \begin{cases} \hat{\alpha}_{e,1,l} & \sim Beta\left(\pi_{e,l} a_{1,l}, \pi_{e,l}(\Sigma_l - a_{1,l})\right) \\ \hat{\alpha}_{e,2,l} & \sim Beta\left(\pi_{e,l} a_{2,l}, \pi_{e,l}(\Sigma_l - a_{2,l})\right) \\ \hat{\Sigma}_{e,l} & \sim Gamma\left(\pi_{e,l}, \frac{\pi_{e,l}}{\Sigma_l}\right) \end{cases}$$

with $\Sigma_l = a_{1,l} + a_{2,l} + a_{3,l}$. The proposed parametrisation for the above beta and gamma distributions are convenient in several respects. First of all, they imply the following properties concerning the expectations of $\hat{\alpha}_{e,1,l}$, $\hat{\alpha}_{e,2,l}$ and $\hat{\Sigma}_{e,l}$:

$$\begin{aligned} E\left(\hat{\alpha}_{e,1,l}\right) &= \frac{a_{1,l}}{\Sigma_l} = E(P_{1,l}) \\ E\left(\hat{\alpha}_{e,2,l}\right) &= \frac{a_{2,l}}{\Sigma_l} = E(P_{2,l}) \\ E\left(\hat{\Sigma}_{e,l}\right) &= \Sigma_l. \end{aligned} \tag{5.27}$$

Moreover, the properties concerning the variances of these quantities, given by

$$
\begin{aligned}
Var\left(\hat{\alpha}_{e,1,l}\right) &= \frac{\frac{a_{1,l}}{\Sigma_l}\left(1-\frac{a_{1,l}}{\Sigma_l}\right)}{\Sigma_l \pi_{e,l}+1} \\
Var\left(\hat{\alpha}_{e,2,l}\right) &= \frac{\frac{a_{2,l}}{\Sigma_l}\left(1-\frac{a_{2,l}}{\Sigma_l}\right)}{\Sigma_l \pi_{e,l}+1} \\
Var\left(\hat{\Sigma}_{e,l}\right) &= \frac{\Sigma_l^2}{\pi_{e,l}}.
\end{aligned} \tag{5.28}
$$

are convenient in two respects. Firstly, they make sure that when describing the uncertainty for

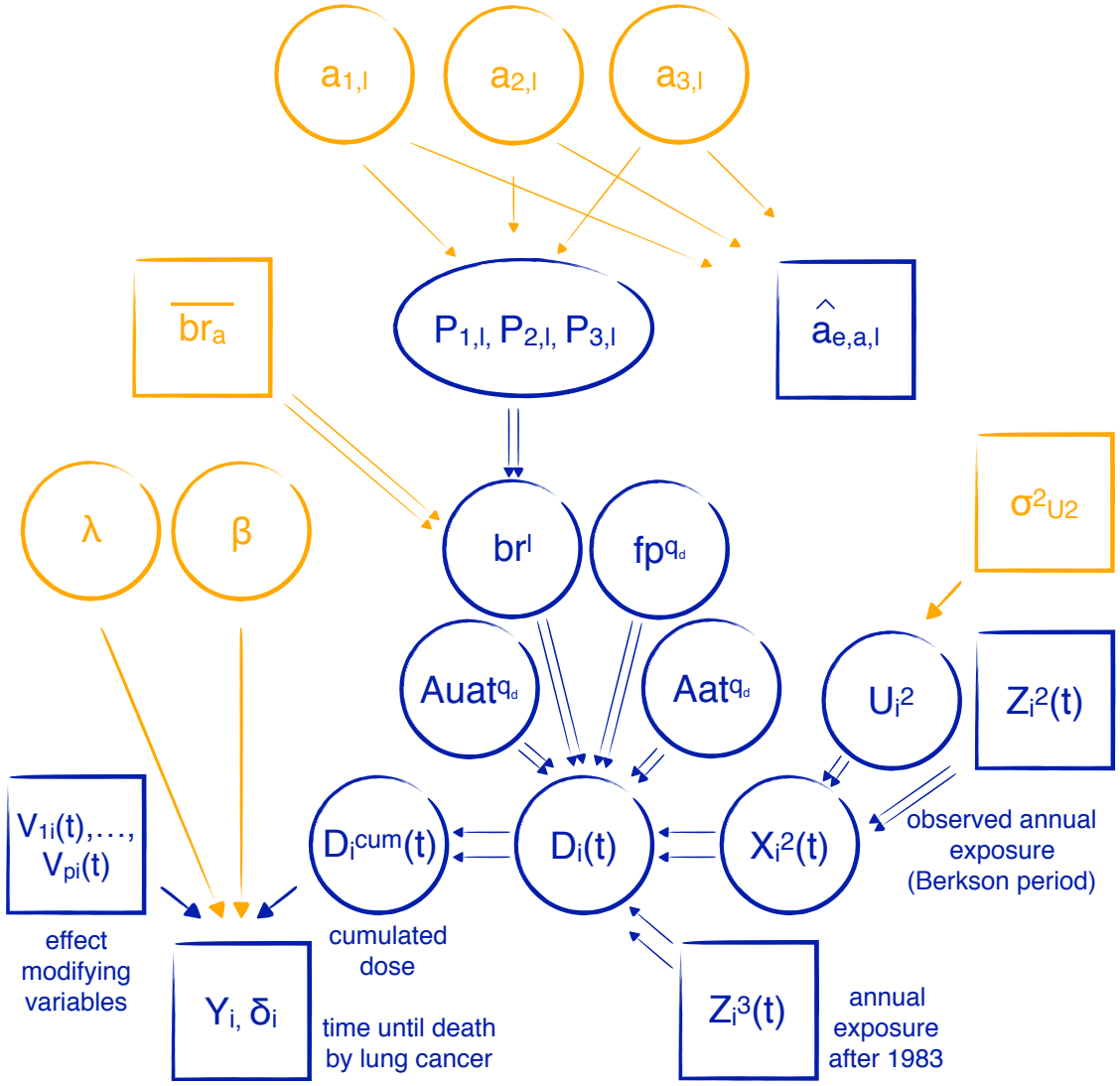

Figure 5.16: DAG when accounting for uncertainty in the input parameters in the dose model, for shared Berkson error and for expert knowledge.

these three quantities, we translate the fact that we can expect more uncertainty in the observed elicited quantities for an expert who indicated low confidence in the information he provided than for an expert who indicated high confidence. Indeed, when $\pi_{e,l}$ increases, the variances for $\hat{\alpha}_{e,1,l}$, $\hat{\alpha}_{e,2,l}$ and $\hat{\Sigma}_{e,l}$ decrease. Thereby, the information provided by confident experts will have more influence on the estimation of $a_{1,l}$ and $a_{2,l}$ and Σ_l than the information given by uncertain

experts. Moreover, we can see that the numerator in the expression of the variances of $\hat{\alpha}_{e,1,l}$ and $\hat{\alpha}_{e,2,l}$ resemble the formula for a random variable V following a Bernoulli distribution with parameter p, i.e., $Var(V) = p(1-p)$[3]. This allows to reflect the fact that we can assume a greater uncertainty in the elicited quantities if the true expected proportion of time a miner spent sitting or in light exercise (given by $E(P_{1,l} = \frac{a_{1,l}}{\Sigma_l})$ and $E(P_{2,l} = \frac{a_{2,l}}{\Sigma_l})$) is close to 0.5 rather than close to 0 or 1.

The unknown parameters in the model $(H1)$ are $a_{1,l}$, $a_{2,l}$ and Σ_l. By setting $a_{3,l} = \Sigma_l - a_{1,l} - a_{2,l}$, we obtain all three parameters of the global Dirichlet distribution f_l^{Supra}.

In a first attempt to combine the information of the three experts via a Supra-Bayesian approach, we assumed the following vague prior distributions for the unknown parameters $a_{1,l}$, $a_{2,l}$ and Σ_l:

$$\Sigma_l \sim Exp(0.01) \tag{5.29}$$

$$a_{1,l}|\Sigma_l \sim Unif(0, \Sigma_l) \tag{5.30}$$

$$a_{2,l}|\Sigma_l, a_{1,l} \sim Unif(0, \Sigma_l - a_{1,l}). \tag{5.31}$$

For Σ_l, we tested the influence of either assuming $\Sigma_l \sim Exp(0.01)$ or a prior distribution that was even more vague with $\Sigma_l \sim Exp(0.001)$.

We conducted Bayesian inference for model (H1) using the prior probability distributions defined above via a MCMC algorithm. Consequently, we obtained G samples of the posterior values $a_{1,l}^{(g)}$, $a_{2,l}^{(g)}$ and $\Sigma_l^{(g)}$ $(g \in \{1, 2, \ldots, G\})$ for the parameters $a_{1,l}$, $a_{2,l}$ and Σ_l. Those posterior samples were used to generate samples if the three-dimensional vector of the proportion a miner in working condition l spent sitting, in light exercise and in heavy exercise

$$(P_{1,l}^{(g)}, P_{2,l}^{(g)}, P_{2,l}^{(g)}) \sim Dirichlet(a_{1,l}^{(g)}, a_{2,l}^{(g)}, (\Sigma_l^{(g)} - a_{1,l}^{(g)} - a_{2,l}^{(g)})). \tag{5.32}$$

The samples of the proportion of time spent sitting, in heavy exercise and in light exercise, in turn, were used to derive samples of the average breathing rate via

$$br^{l,(g)} = P_{1,l}^{(g)} \cdot \bar{b}r_1 + P_{2,l}^{(g)} \cdot \bar{b}r_2 + P_{3,l}^{(g)} \cdot \bar{b}r_3. \tag{5.33}$$

These posterior samples of average breathing rate in working condition l are considered to be samples of the elicited prior distribution on this uncertain input parameter in the dose model. In the context of the Supra-Bayesian approach for the aggregation of expert knowledge, it is interesting to note that we can consider the probability model $(H1)$ presented in this section as an additional sub-model of the hierarchical structure presented in section 5.1.4. In this framework, the dose model is defined as

$$D_i(t) = \frac{br^l}{cbr} \cdot (fp^{q_d} \cdot cunat^{q_d} + (1 - fp^{q_d}) \cdot cat^{q_d}) \cdot X_i^q(t) \tag{5.34}$$

where $br_i(t) = br^l$ for all miners i and all exposure years t corresponding to the same working condition $l \in \{1, 2, 3, 4, 5, 6\}$. Moreover $fp_i(t) = fp^{q_d}$, $cunat_i(t) = cunat^{q_d}$ and $cat_i(t) = cat^{q_d}$ for all miners i and all exposure years t corresponding to the same dose period q_d (i.e. $q_d = 1$ for years $t \leq 1977$ and $q_d = 2$ for years $t > 1977$). The relation between all observed and unobserved quantities when accounting for shared Berkson error, dose uncertainty and the information provided by the three experts can be described by the DAG presented in Figure 5.16.

[3]The function $p(1-p)$ with is concave for values of $p \in [0, 1]$ with maximum value at 0.5

5.3 Bayesian inference and model checking

5.3.1 Implementing a MCMC algorithm to conduct Bayesian inference for the proposed hierarchical models

In order to conduct Bayesian inference for the full hierarchical models proposed in this work to account for exposure and dose uncertainty through the combination of the disease, the measurement, the exposure and the dose model described in section 5.1, we were faced with a number of challenges.

Firstly, as described in section 4.4, when accounting for measurement error, we treat true and unknown exposure $X_i(t)$ of miner i at time t as a latent variable. In the context of a MCMC algorithm, this high-dimensional unknown quantity has to be updated at each iteration. With 5086 miners and on average 11.8 exposure years per miner (see section 2.2.2), there were about 50000 true exposure values to be updated at each iteration of a MCMC algorithm when accounting for unshared measurement error in the full cohort. The updating of high-dimensional unknown quantities can pose serious convergence issues and lead to very long time in computation in MCMC sampling [298].

On the other hand, the implementation of statistical inference for the hierarchical model presented in section 5.1 requires a very flexible algorithm. In particular, the fact that lung cancer mortality is linked to cumulated exposure or cumulated dose introduces a certain complexity in the problem. As a time-varying variable, cumulated exposure of miner i at time t, $X_i^{\text{cum}}(t)$, is defined as the sum of the true exposure values miner i received until time t. This implies that when accounting for exposure uncertainty, we not only need to update latent exposure $X_i(t)$, but we also need to compute the corresponding values for $X_i^{\text{cum}}(t)$ at each iteration. This operation can be very time-consuming. Indeed, while the calculation of the cumulated sum of the vector of exposure values for each individual i may appear simple, it is complicated by the fact that each miner may have received an arbitrary number of exposure values during his working career. Therefore, there is no straightforward operation that could be applied to the whole vector of exposure values to achieve the cumulation of $X_i(t)$ in a time-efficient manner. The use of a for-loop to achieve this cumulation for each miner independently cannot be considered a satisfactory solution as it would be too time-consuming in light of the large number of iterations which are necessary to explore a very high-dimensional posterior distribution. Finally, the consideration of time since exposure as effect-modifying variable implies a further complexity, because at each iteration, we have to calculate three different sums of the exposure values that miner i received in different exposure periods until time t to determine $X_{i,5-14}^{\text{cum}}(t)$, $X_{i,15-24}^{\text{cum}}(t)$ and $X_{i,25+}(t)$.

In principle, an efficient alternative to MCMC sampling in the situation of high-dimensional latent variables would be to conduct approximate Bayesian inference, for instance via Integrated Nested Laplace Approximation (INLA) [415, 191, 396, 416]. INLA is an approximate method of Bayesian inference for latent Gaussian models, which was first proposed by Rue et al. (2009) [415] in the context of generalised linear models. Martino et al (2011) [396] generalised this method to proportional hazards models and Muff et al. (2015) illustrated the usefulness of INLA when the aim is to account for measurement error in explanatory variables [191]. However, as pointed out by Roberts (2009) in his discussion on the paper of Rue et al. (2009) [415], the calibration of Laplace approximations appears to require a high level of expertise. Although INLA can be

used as a "black box" [396] in its R implementation for instance, this implementation potentially lacks the flexibility to account for multiple high-dimensional latent variables. Moreover, it is questionable whether we could have adequately described all uncertain input parameters in the dose model as Gaussian variables. For the unattached fraction $fp_i^{q_d}$ of miner i in dose period $q_d \in \{1, 2\}$ and the activity median diameter of the attached radon progeny $Aat_i^{q_d}$, which are supposed to follow a log-normal distribution (see section 5.1.5) this could easily be achieved by a log-transformation. The activity median diameter of the unattached radon progeny $Auat_i^{q_d}$, on the other hand, is supposed to follow a uniform distribution. Similarly, when accounting for expert knowledge to derive an informative prior probability distribution assigned to breathing rate br^l, br^l is defined as a function of the three components of a Dirichlet distribution. Furthermore, the treatment of the statistical dependence between the true and latent exposure $X_i^q(t)$ in the case of shared Berkson error might have implied substantial extensions of INLA (see Muff et al. (2015) [191] for a discussion of the possibility of these extensions as well as concerning the possibility of the treatment of non-Gaussian latent variables) and it was not in the scope of this work to develop these extensions. Finally, our aim was to present a general algorithm that could flexibly be extended to various situations and potentially be of use to other researchers, who seek to account for multiple sources of uncertainty in epidemiological studies. Based on these arguments, we chose to conduct Bayesian inference via a Metropolis-within-Gibbs algorithm. The Metropolis-Hastings algorithm remains the most universal and flexible way of conducting Bayesian inference in spite of its problems in high dimensions (see section 4.2.3).

In light of the challenges mentioned in the beginning of this section, it proved infeasible to conduct Bayesian inference for the different hierarchical models proposed in this work in an available software package, for instance via WinBUGS or Jags. We therefore chose to implement an MCMC algorithm to conduct Bayesian inference on the different proposed hierarchical models in Python 2.7. To exploit the full flexibility of Python, we chose an object-oriented implementation. The class diagram describing the algorithm is given in Figure 5.17. In an object-oriented implementation, we can flexibly create independent modules for the different tasks of the algorithm. Moreover, it allows to reuse part of the code for other applications. For instance, if we chose to modify the calculation of detriment-weighted absorbed lung dose $D_i(t)$ by using a more complex dose model than the one we presented in section 5.1.4, we would merely have to change the method *calculate_dose* of the class LatentVariable (see Figure 5.17), which is responsible for dose calculation. Moreover, when implementing an iterative algorithm there are many parameters which will be the same for each iteration. It is convenient to use an object-oriented architecture in this case, because these parameters can be treated as attributes of the different classes. For instance, the MCMC class in the class diagram in Figure 5.17 has the attributes $measurement\_error, shared\_measurement\_error, dose\_uncertainty$, which are boolean variables indicating whether to account for unshared measurement error, shared measurement error and dose uncertainty, respectively. These attributes will stay the same for each iteration of the MCMC algorithm. These attributes can directly be accounted for in the initialisation of the MCMC class to define methods which are specific to the accounting for dose uncertainty and measurement error. Moreover, an object-oriented implementation can avoid both the use of numerous if and else statements in functions to handle the different conditions and the specification of a large number of parameters in the definition of functions. The attribute *path* (referring

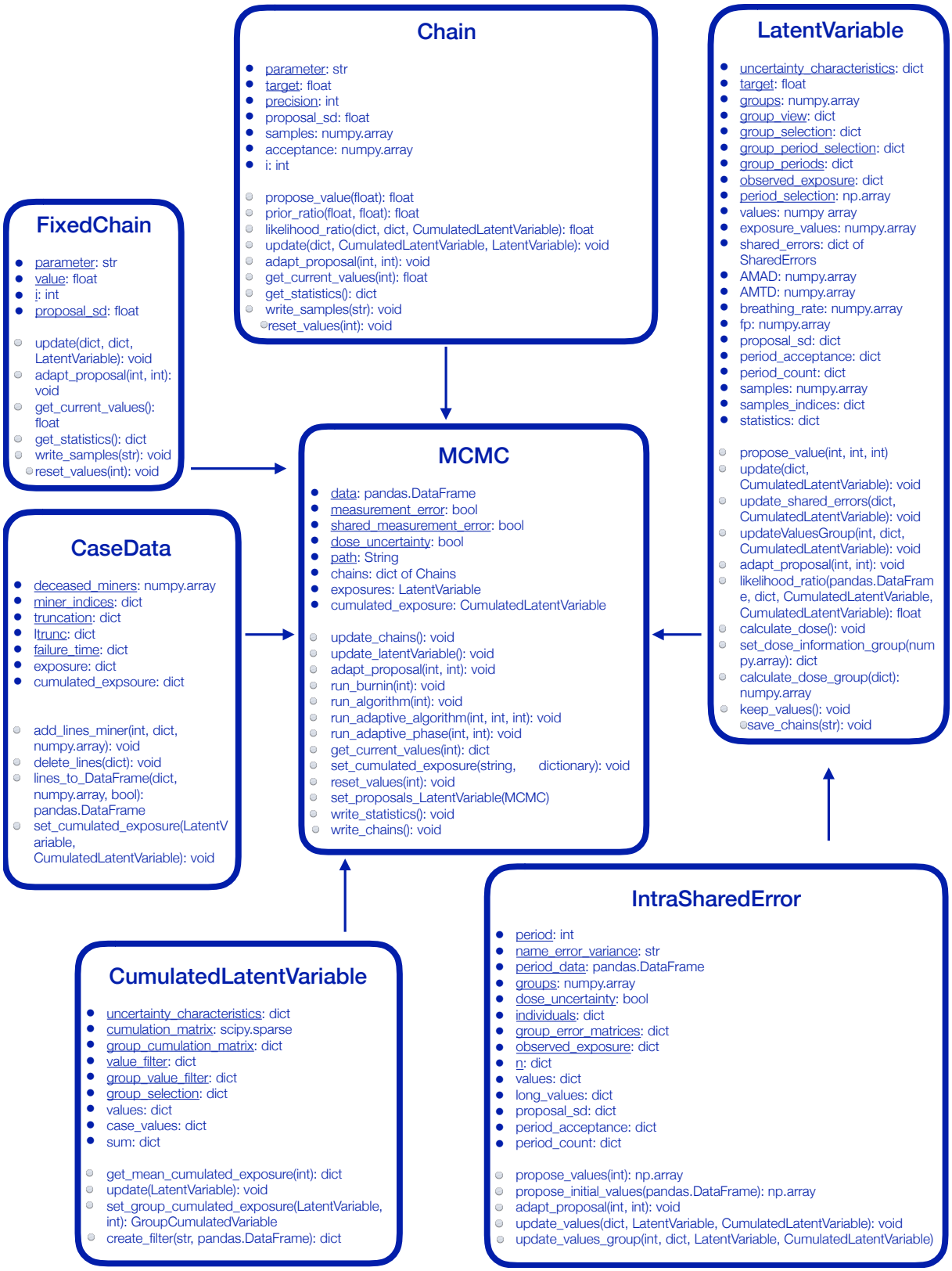

Figure 5.17: The class diagram of the implemented Markov Chain Monte Carlo algorithm.

to the path at which the output files should be stored) of the MCMC object, for instance, can easily be accessed by $self.path$ in the methods of the MCMC object and this parameter therefore does not have to be specified in any calls of the MCMC method. Thereby, object-oriented code is both more readable and easier to modify.

With the exception of the parameters in the exposure model, the full conditional distributions of the unknown quantities that had to be updated when conducting Bayesian inference for the proposed hierarchical models were intractable. Therefore, we employed a Metropolis-within-Gibbs algorithm to update the unknown quantities, where only the parameters of the exposure model were updated via Gibbs sampling steps. We chose a component-wise scheme, updating each dimension independently. The only exceptions were the update of the high-dimensional vector of true and latent exposure values $X_i(t)$ quantities when accounting for measurement error or dose uncertainty and the three parameter $a_{1,l}$ and $a_{2,l}$ and Σ_l of model $H1$ proposed for the elicitation of prior knowledge on average breathing rate via a Supra-Bayesian approach, which we updated simultaneously (see section 5.2.3). For the update of the high-dimensional vector of exposure values, we used block-wise sampling, updating all exposure values for a specific calendar period and a homogenous group of miners in a single step. Updating all true and unknown exposure values at once would have led to a small acceptance rate of the Metropolis-Hastings algorithm while a component-wise version may have taken too much time in computation.

Homogeneous groups of miners were determined via a hierarchical ascendent clustering algorithm using variables such as principal type of mine, principal location, the calendar periods during which the miner was exposed and principal type of work of miners. In order to use categorical variables in this clustering algorithm, we first performed a multiple correspondence analysis to subsequently apply the clustering algorithm on the resulting factor values. The clustering algorithm yielded 50 latent exposure vectors of varying sizes. In order to avoid the simultaneous updating of a large number of exposure values at once, we decided to split the obtained clusters until there were at most 150 unknown exposure values per cluster and per period. While it could have been possible to use this constraint directly in the clustering algorithm [417, 418], it was not in the scope of this work to implement a constrained clustering algorithm for this purpose. We used the same groups for the updating of the high-dimensional quantities intervening in the dose model and for the updating of the measurement error components U_i^1 and U_i^2 when accounting for shared Berkson error. As mentioned in the beginning of this section, the cumulation of latent exposure $X_i(t)$ to obtain cumulated exposure $X_i^{\text{cum}}(t)$ at each MCMC iteration presented a major challenge in the development of the algorithm. In principle, this cumulation could be achieved via for-loops where for each miner, we determine the cumulated sum, for instance via the function *cumsum* in the *numpy* package in Python. Unfortunately, the use of for-loops in languages like R or Python is very time consuming, in contrast to Java or C++ for instance. However, both R and Python are very efficient when it comes to vector operations. We exploited this fact by basing the cumulation of latent exposure or dose values on matrix multiplication. For instance, if we imagine three miners where the first miner received the five exposure values $X_{11}, X_{12}, X_{13}, X_{14}$ and X_{15} during his working career, miner two received two exposure values X_{21} and X_{22} and miner 3 received the three exposure values X_{31}, X_{32} and X_{33}, we can achieve the cumulation of the exposure values by the following

matrix multiplication:

$$\begin{pmatrix} X_{11} \\ X_{11}+X_{12} \\ X_{11}+X_{12}+X_{13} \\ X_{11}+X_{12}+X_{13}+X_{14} \\ X_{11}+X_{12}+X_{13}+X_{14}+X_{15} \\ X_{21} \\ X_{21}+X_{22} \\ X_{31} \\ X_{31}+X_{32} \\ X_{31}+X_{32}+X_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \\ X_{14} \\ X_{15} \\ X_{21} \\ X_{22} \\ X_{31} \\ X_{32} \\ X_{33} \end{pmatrix} \tag{5.35}$$

As mentioned in section 5.1.1, the size of the data frames to reflect the full exposure history of the miners in the French cohort of uranium miners varied between 40162 rows for the post-55 cohort and 119073 rows when accounting for the effect-modification by time since exposure in the full cohort. To be able to implement the matrix multiplication based on square matrices with dimension of up to 119073×119073, we used the sparse module in the *scipy* package in Python. This module uses the fact that the cumulation matrices we created are sparse, i.e. most of the elements of the matrices are zero, to reduce the amount of memory required to store them. Moreover, the use of this module accelerated all matrix operations by accounting for the fact that the matrices are sparse. When compared to an implementation based on for loops, the implementation using sparse matrices was about 15 times faster in Python 2.7, thereby implying a considerable acceleration in the MCMC algorithm.

We also relied on the use of sparse matrices when updating shared components of exposure and dose uncertainty. When considering the updating of shared Berkson error components, our aim is to update the true exposure values under the measurement model $X_i(t) = Z_i(t) \cdot U_i$. In other words, to account for shared exposure uncertainty, we have to update all true and unknown exposure values while respecting the constraint that the error component of each miner is shared for all years of exposure in a given exposure period. In our example involving three miners, we can imagine that all exposure values were received in the same period. In this case, we have to obtain an error vector $U_i = (U_1, U_1, U_1, U_1, U_1, U_2, U_2, U_3, U_3, U_3)$ which can be multiplied by the observed exposure values $Z_i(t)$ in order to obtain true exposure $X_i(t)$ while respecting the constraint that the error component for a miner is shared for all years of exposure. In theory, this could be implemented by selecting the corresponding values in a for loop in the following way: for each miner in the group of miners, we multiply all the observed exposure values received in a certain period of exposure with the same error component U_i. In our simple example with three miners and ten exposure values, it would be acceptable to do this operation a large number of times (e.g. for 100.000 iterations for instance). However, we have to update several thousand error components for the miners that were exposed in the first and in the second exposure period and to repeat the same for-loop during 100.000 iterations would be prohibitively time consuming in this situation. A more efficient way to realise this multiplication in Python is again to use sparse matrices, similar to the cumulation of exposure values. In the context of this sparse matrix implementation, we constructed a matrix for each group and each period that

could achieve the following matrix implementation:

$$\begin{pmatrix} U_1 \\ U_1 \\ U_1 \\ U_1 \\ U_1 \\ U_2 \\ U_2 \\ U_3 \\ U_3 \\ U_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} \tag{5.36}$$

When comparing the implementation based on sparse matrices with an implementation that achieves the same operation via a for loop, the sparse matrix implementation was about 20 times faster than the for-loop implementation. As this operation has to be done at every iteration of the algorithm, the implementation based on sparse matrices therefore leads to a substantial gain in computation time of the MCMC algorithm.

To further accelerate the algorithm, we implemented accelerated evaluations of the measurement and the exposure model based on the fact that basic operation like additions are less time-consuming than divisions or the application of the exponential and the logarithmic function. The demonstration of the validity of these accelerated evaluations can be found in Appendix A. When the updating unknown quantities via a Metropolis-Hastings step, we either chose normal or log-normal random walks as proposal distributions for all unknown parameters $\theta_l, \forall l \in 1, \ldots, p$ with p the number of unknown parameters to update, yielding either $\theta_l^{\text{cand}} | \theta_l^{(t)} \sim \mathcal{N}(\theta_l^{(t)}, \sigma_{\theta_l}^2)$ or $\theta_l^{\text{cand}} | \theta_l^{(t)} \sim \mathcal{LN}(\log(\theta_l^{(t)}), \sigma_{\theta_l}^2)$ for each iteration t, where $\theta_l^{(t)}$ is the current value of the parameter θ_l at iteration t and θ_l^{cand} is the corresponding candidate value. The only exception was the proposal distribution for the activity size distribution of the unattached radon progeny $Auat_i(t)$ defined in dose model A_1 (section 5.1.4). For this high-dimensional input parameter in the dose model we implemented an independence sampler, in which candidate values were sampled in a uniform distribution between 0.5 nm and 1.5 nm, irrespective of the current values of the Markov chain. Based on the literature, we supposed that this input parameter could only take values between 0.5 nm and 1.5 nm (see section 5.1.5). Moreover, preliminary MCMC runs showed that the values of this input parameter were only poorly informed by the disease model. In light of these facts, it seemed that an independence sampler was an efficient way to explore the space of possible values. Since the other high-dimensional unknown quantities intervening in the dose model could only take positive values, we opted for a log-normal random walk algorithm to update these quantities [419]. The variances of the proposal distributions $\sigma_{\theta_l}^2$ were calibrated for each parameter θ_l so as to obtain acceptance rates close to 0.2 for the high-dimensional unknown quantities and close to 0.4 for the unknown parameters. As it was impossible to fine-tune the proposal variances per hand, we chose an adaptive version of the Metropolis-Hastings steps that preceded the burn-in phase.

The development of the algorithm was largely conducted using a test-driven development approach [420, 421], in which each function was tested directly after its development by several unit tests, as this approach is advocated as being more efficient and less error-prone than classical

programming approaches [422, 423, 424]. Unit tests were complemented with test on simulated data to check the global performance of the algorithm. The results of a simulation study concerning the performance of the implemented Bayesian hierarchical approach to account for unshared measurement error can be found in Appendix B. We used Mercurial version 4.3.1 as source control management tool. Trace plots and the Gelman Rubin statistic [306] were used to check the convergence of the Markov chains.

5.3.2 Model comparison and model checking

Model comparison

Competing disease models were compared via the Deviance Information Criterion (DIC) [425, 426]. DIC values are a measure for the fitting abilities of a model with smaller DIC values indicating a better fit to the data. The DIC can be seen as a penalised likelihood criterion:

$$DIC = D_m(y) + p_D. \tag{5.37}$$

Let $D(y, \boldsymbol{\theta}) = -2\log([Y|\boldsymbol{\theta}])$ be the deviance that measures the fitting abilities of the probability model with parameters $\boldsymbol{\theta}$ to the outcome Y. $D_m(y)$ is the posterior expectation of this deviance, i.e. $D_m(y) = E(D(y, \boldsymbol{\theta})$ which can be estimated by $\frac{1}{K}\sum_{k=1}^{K} D(y, \boldsymbol{\theta}^{(k)})$ with K the sample size of the posterior on $\boldsymbol{\theta}$. p_D can be interpreted as the "effective number of parameters" or the model complexity, thereby penalising more complex models. $p_D = D_m(y) - D(y, E(\boldsymbol{\theta}|y))$ where $D(y, E(\boldsymbol{\theta}|y))$ is the deviance of the posterior expectation of $\boldsymbol{\theta}$ that can be estimated by $D(y, \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\theta}_k)$. The smaller this criterion, the better are the fitting abilities of the associated model.

Testing the proportional hazard assumption

The disease model

$$h_i(t; \boldsymbol{\theta}) = h_0(t) \cdot \varphi(X_i^{cum}(t), V_{i1}(t), \ldots, V_{ip}(t), \boldsymbol{\theta}) \tag{5.38}$$

we introduced in section 5.1.1 relies critically on the assumption that for two individuals i and i' the ratio between their instantaneous hazard rates only depends on time via the values of their time-varying covariates, i.e. $\frac{h_i(t;\boldsymbol{\theta})}{h_{i'}(t;\boldsymbol{\theta})} = \frac{\varphi(X_i^{cum}(t), V_{i1}(t), \ldots, V_{ip}(t), \boldsymbol{\theta})}{\varphi(X_{i'}^{cum}(t), V_{i'1}(t), \ldots, V_{i'p}(t), \boldsymbol{\theta})}$. For the linear EHR model without effect modification ($\mathcal{D}_1$) this reduces to $\frac{h_i(t;\boldsymbol{\theta})}{h_{i'}(t;\boldsymbol{\theta})} = \frac{1+\beta X_i^{cum}(t)}{1+\beta X_{i'}^{cum}(t)}$. This assumption is commonly referred to as the proportional hazards assumption. We tested this assumption via Schoenfeld residuals and the Harrel test for the linear Cox model ($\mathcal{D}_2$) in R.

5.4 Studying the effects of measurement model misspecification on simulated data

We designed two simulation studies based on the exposure data of the French cohort of uranium miners (see section 2.2.2) to assess the effects of measurement model misspecification when conducting statistical inference for proportional hazards models. More precisely, our aim was to assess the impact of different types of measurement error on risk estimation and on the exposure-response relationship when measurement error is not accounted for. In the context of

these simulation studies, we considered different measurement models to generate error-prone exposure data, which we will denote as $\mathcal{M}_{Sk}$ for the k$^{\text{th}}$ measurement model in simulation study 2 and different disease models to generate failure times, which we will denote as $\mathcal{D}_{Sk}$ for the k$^{\text{th}}$ disease model. Statistical inference on the generated exposure data was always conducted without accounting for measurement error. In other words, the only measurement model we assumed when conducting statistical inference in both simulation studies was the model

$$\mathcal{M}_{E0} : X_{ij}(t) = Z_{ij}(t) \ \forall i \ \forall t \tag{5.39}$$

where $X_{ij}(t)$ and $Z_{ij}(t)$ denote the true and observed exposure value received by miner i at time t, assuming that worker i belonged to group j. We considered these groups in the simulation study in order to be able to generate measurement error shared between workers belonging to a same group (see section 3.1.1). As the French cohort of uranium miners did not present a natural partition into groups of workers, we used the homogeneous groups of workers that had been created via a hierarchical ascendant clustering algorithm based on covariates concerning the following job characteristics: principal type of mine, principal location and principal type of job (see section 5.3.1). Similar to the measurement models assumed in the analysis of the observed failure times in the French cohort of uranium miners (see section 5.1.2), we assumed unbiased, multiplicative and log-normal measurement error with mean $\mu = -\frac{\sigma^2}{2}$ and variance σ^2 of the normal distribution describing the log-transformed errors. In both simulation studies, we generated failure times according to the two proportional hazards models $\mathcal{D}_{S1}$ (linear EHR model without effect modification) and $\mathcal{D}_{S2}$ (the Cox model) described in section 5.1.1. $\mathcal{D}_{S1}$ and $\mathcal{D}_{S2}$ are exactly the same models as the linear EHR model $\mathcal{D}_1$ and the linear Cox model $\mathcal{D}_2$ that we presented in section 5.1.1. Additionally, we compared the effects of measurement error on statistical inference for moderate and large effect sizes of radon exposure on lung cancer mortality. We chose the value $\beta = 2$ as moderate effect size and $\beta = 5$ as a large effect.

5.4.1 Simulation study 1: The impact of shared and unshared measurement error on risk estimation

In simulation study 1, the measurement models that we used to generate the error-prone exposure data described an error structure that was assumed to be the same for all exposure periods. More complex measurement models that assume that the error structure varies over time will be treated in simulation study 2, described in section 5.4.2. Several authors have argued that error components shared between individuals might have fundamentally different consequences on statistical inference than unshared measurement errors [135, 118, 119, 138, 153, 427]. To our knowledge, there are no simulation studies confirming this assertion for proportional hazards models, which possibly presents the most widely applied class of models in medical research. The aim of simulation study 1 was therefore to highlight and compare the effects of shared and unshared exposure measurement error on risk estimation in the context of proportional hazards models. To that purpose, we generated error-prone exposure data by varying the following characteristics:

1. *Type of sharing*

   We compared the effects of unshared error components, error components shared between

workers, error components shared for several exposure years of the same worker (i.e. within workers) and error components that were both shared between workers and within workers (see section 3.1.1).

2. *Type of error*
   We assumed either Berkson or classical measurement error.

The combination of these characteristics resulted in 8 different measurement models for the generation of error-prone exposure data. As we compared the effects of measurement error on two different disease models, these choices resulted in 16 distinct simulation models to study the impact of measurement error on risk estimation when measurement error is not accounted for in statistical inference (i.e. under measurement model $\mathcal{M}_{E0}$). Under each of these 16 simulation models, four scenarios corresponding to two different values of the following parameters were considered:

1. *Effect size*
   We compared large ($\beta = 5$) and moderate ($\beta = 2$) effect sizes in the generation of failure times.

2. *Measurement error variance*
   We assumed either a large ($\sigma_U^2 = 0.9$) or a moderate ($\sigma_U^2 = 0.1$) value for the variance of the log-transformed measurement errors.

For each case, we generated 100 data sets according to the approach described in section 5.4.3. We conducted statistical inference via the MCMC algorithm presented in 5.3.1. As we did not account for measurement error in the estimation of risk parameters, we only conducted inference on the disease model which had been used to generate failure times. In other words, we fitted the EHR model $\mathcal{D}_{E1}$ for data generated according to $\mathcal{D}_{S1}$ and the Cox model $\mathcal{D}_{E2}$ for data sets generated according model $\mathcal{D}_{S2}$. Note that we chose large variances (1000) for the normal prior assigned to the risk coefficient β.

5.4.2 Simulation study 2: The effects of measurement error characteristics on the shape of the exposure-response curve

Changes in the methods of exposure assessment in occupational cohort studies can create complex patterns of exposure uncertainty, where the type and magnitude of measurement error can vary over time. It has been suggested that the fact that exposure uncertainty and the magnitude of exposure are both highest for the earliest exposure periods may cause an attenuation of the exposure-response curve for high exposure values, a phenomenon which is frequently observed in occupational cohort studies [189, 17, 14]. Stayner et al. (2003) [17] and Steenland et al. (2015) [14] examined the effects of heteroscedastic measurement error on the shape of the exposure-response curve and only found a modest attenuation of the exposure-response curve at high exposure values. However, the authors treated cumulated exposure in an occupational cohort as a time-fixed variable known at study entry, thereby ignoring both its time-varying nature and the possibility of exposure uncertainty components shared within individuals.
We therefore designed a second simulation study, which will be referred to as simulation study 2

in the following, to reassess this finding when explicitly accounting for the time-varying nature of cumulated exposure and when modelling measurement error on its natural level of occurrence, i.e. on the annual exposure values instead of modelling measurement error on the sum of these values. The aim of simulation study 2 was therefore to quantify the effects of different error structures on the observed shape of the exposure-response curve when conducting statistical inference that does not account for these error structures (i.e. under $\mathcal{M}_{E0}$). In this simulation study, we assumed three different exposure periods to reflect the changes in the method of exposure assessment in the French cohort of uranium miners: 1946-1955, characterised by a retrospective exposure estimation by experts, 1956-1982, characterised by a method of prospective and group-based exposure assessment based on ambient measurements and 1983-2007, characterised by personal dosimetry (see section 2.2.3 for more details on the methods of exposure assessment in the French cohort of uranium miners). We compared the estimated shape of the exposure-response curve for exposure data generated without measurement error (i.e., according to model $\mathcal{M}_{S0}$) and when generated with unshared and homoscedastic Berkson error for the three exposure period, denoted $\mathcal{M}_{S1}$, with more complex and more realistic measurement models. In particular, we compared the results for $\mathcal{M}_{S0}$ and $\mathcal{M}_{S1}$ with three measurement models $\mathcal{M}_{S2}$, $\mathcal{M}_{S3}$ and $\mathcal{M}_{S4}$:

1. $\mathcal{M}_{S2}$ assumed unshared exposure measurement error for the three exposure periods. Moreover, $\mathcal{M}_{S2}$ described errors occurring in the first and in the second exposure period as Berkson error, while the error occurring in the third exposure periods were described as classical measurement error. We assumed the variances $\sigma^2_{U,1} = 0.9$, $\sigma^2_{U,2} = 0.15$ and $\sigma^2_{U,3} = 0.01$ for the log-transformed errors occurring in the first, the second and the third period, respectively. These values were chosen in accordance with the characterisation of exposure uncertainty by Allodji et al. (2012) [33, 34].

2. In contrast to model $\mathcal{M}_{S2}$, measurement model $\mathcal{M}_{S3}$ assumed a more complex error structure in the three exposure periods by accounting for individual job conditions and worker practices in the cohort, similarly to measurement model $\mathcal{M}_2$ (described in section 5.1.2). In line with the assumption that individual worker characteristics can lead to Berkson error shared for several years of exposure of the same miner when a strategy of group-level exposure assessment is employed, measurement model $\mathcal{M}_{S3}$ assumed a combination of unshared Berkson error and Berkson error shared within workers for the first and the second exposure period with variances $\sigma^2_{U,1} = 0.09$ and $\sigma^2_{U,2} = 0.03$ for unshared Berkson error and $\sigma^2_{U*,1} = 0.81$ and $\sigma^2_{U*,2} = 0.12$ for shared Berkson error. As the third exposure period between 1983 and 2007 was characterised by personal dosimetry, there is no reason to assume a shared error component due to individual worker characteristics in this period (see section 5.1.2). As in model $\mathcal{M}_{S2}$, we therefore assumed unshared classical measurement error for the third exposure period with $\sigma^2_{U,3} = 0.01$.

3. Finally, the imprecision of the measurement device used in group-level exposure estimation can lead to a component of classical measurement error, which is shared between a group of workers. For instance, in the retrospective exposure reconstruction for the exposure years between 1946 and 1955 in the French cohort of uranium miners, a group of experts

was asked in 1980s to estimate the average monthly exposure value for a mining location. As the estimation for several years of the same mining location was based on the same hypotheses and measurements, the estimated exposure levels were often the same for several exposure years of the same mine. The estimated monthly exposure level in a mine was then used to determine the exposure values for all years of exposure that a miner spent in the mining location. In other words, the measurement error occurring in the estimation of the monthly exposure level in a mine affected all exposure years of all miners employed in a certain mine simultaneously. We can therefore consider that the retrospective exposure reconstruction in the first exposure period of the French cohort of uranium miners led to a classical error component that was shared both between miners and within miners. For the exposure period between 1956 and 1982, based on a prospective method of group-exposure assessment via ambient measurements, on the other hand, the precision on the measurement device might be shared between miners but not for several exposure years of the same miner. Measurement model $\mathcal{M}_{S4}$ accounted for these shared classical error components due to the imprecision of the measurement device in the following way: It assumed a combination of a classical error component that was shared both between and within miners for the first exposure period ($\sigma^2_{U*,1} = 0.81$) and an unshared Berkson component ($\sigma^2_{U,2} = 0.09$). Similarly, in the second exposure period, it assumed a combination of a classical error component that was shared between workers ($\sigma^2_{U*,2} = 0.12$) and an unshared Berkson component ($\sigma^2_{U,2} = 0.03$). Finally, in the same way as model $\mathcal{M}_{S2}$ and $\mathcal{M}_{S3}$, it assumed unshared classical measurement error for the last exposure period with $\sigma^2_{U,3} = 0.01$.

Table 5.2: Summary of the measurement models used in the generation of error-prone exposure data in simulation study 2

Measurement model	Description
$\mathcal{M}_{S0}$	No measurement error in all three exposure periods
$\mathcal{M}_{S1}$	Unshared and homoscedastic Berkson error in all three exposure periods
$\mathcal{M}_{S2}$	Unshared heteroscedastic Berkson error in exposure periods 1 and 2 Unshared classical measurement error in exposure period 3
$\mathcal{M}_{S3}$	Combination of a component of unshared Berkson error and a component of Berkson error shared within workers in exposure periods 1 and 2 Unshared classical measurement error in exposure period 3
$\mathcal{M}_{S4}$	Combination of a component of unshared Berkson error and a component of classical measurement error shared both within and between workers in exposure period1 Combination of a component of unshared Berkson error and a component of classical measurement error shared between workers in exposure period2 Unshared classical measurement error in exposure period three

In order to obtain an error variance for model $\mathcal{M}_{S1}$, which was comparable to the global error variance for the three models $\mathcal{M}_{S2}$, $\mathcal{M}_{S3}$, $\mathcal{M}_{S4}$ on the log-scale, we chose $\sigma^2_{U,q} = 0.2$ for the three exposure periods $q \in \{1, 2, 3\}$ for $\mathcal{M}_{S1}$. Table 5.2 summarises the simulation measurement models that were used in simulation study 2. We generated 100 data sets for each scenario. To investigate the possibility of measurement error to induce a non-linear exposure-response relationship, we estimated parameter values in an EHR ($\mathcal{D}_{E3}$) and a Cox model ($\mathcal{D}_{E4}$) based on natural cubic splines. We chose interior knots at the 20th, 40th, 60th and 80th percentile of the exposure distribution of cases, i.e. miners who died of lung cancer in our simulation study.
While the fitting of the disease models $\mathcal{D}_{E3}$ and $\mathcal{D}_{E4}$, based on cubic splines, allows for a graphical evaluation of the impact of different measurement error characteristics, the resulting parameter estimates in these models are not easily interpretable. Consequently, we complemented these findings with results on continuous piecewise-linear models with a breakpoint at 100 WLM assuming an EHR structure ($\mathcal{D}_{E5}$) and a Cox model ($\mathcal{D}_{E6}$) to obtain slope estimates for high and low exposure values. Finally, to study the effects of different measurement models on the choice of the disease model, we conducted statistical inference for both models $\mathcal{D}_{E1}$ and $\mathcal{D}_{E2}$ for all data sets in simulation study 2, regardless of the disease model that was used for data generation. Table 5.3 summarises the six disease models that were used to conduct statistical inference in simulation study 2.

Table 5.3: Summary of the disease models used in statistical inference for the data sets generated in simulation study 2

Disease model	Description
$\mathcal{D}_{E1}$	Linear Excess Hazard Ratio (EHR) model
$\mathcal{D}_{E2}$	Linear Cox model
$\mathcal{D}_{E3}$	EHR model based on cubic splines
$\mathcal{D}_{E4}$	Cox model based on cubic splines
$\mathcal{D}_{E5}$	Piecewise-linear EHR model
$\mathcal{D}_{E6}$	Piecewise-linear Cox model based on cubic splines

5.4.3 Data generation

The outcome of interest for the two simulation studies was time until death by lung cancer of miner i, which can be described by the couple (Y_i, δ_i) as described in section 5.1.1. This outcome depends on the true exposure history of miner i. As described in section 5.1.1, the exposure history of a worker is best described by the sum of the exposure values received until time t, $X_i^{\text{cum}}(t)$. Considering cumulated exposure as time-varying variable allows to model exposure measurement error on its natural level of occurrence, namely on the observed exposure

values $Z_i(t)$ corresponding to true exposure $X_i(t)$ instead of modelling measurement error on the sum of the exposure values that miner i received during his working career. It is generally acknowledged that the generation of failure times as a function of time-varying covariates is very challenging [31, 428, 429, 430]. In this situation, it is impossible to adopt the standard approach for the generation of failure times, which consists in the inversion of survival functions [31], as this would imply inverting the expression $-\int_0^t h_0(u)\varphi(X_i^{\text{cum}}(u), V_{i1}(u), ..., V_{ip}(u), \boldsymbol{\theta})du$, which is infeasible when the changes in the time-varying explanatory variables cannot be modelled by a parametric function [431, 428]. Faced with this difficulty, Sylvestre et al. (2008) [428] advocate a permutational algorithm which incorporates a rejection sampler, but this method is far less computationally efficient than the simple inversion of survival functions. We therefore adapted a method for the generation of failure times as a function of time-varying explanatory variables proposed by Zhou (2001) [432] and further developed by Hendry (2014) [430], which is based on the generation of piecewise-exponential variables.

Concerning the generation of error-prone exposure data, we generated Berkson error by setting the values of observed exposure $Z_i(t)$ equal to the exposure values observed in the cohort to obtain true exposure $X_i(t)$ by multiplying $Z_i(t)$ with a measurement error term. Conversely, when generating classical measurement error, we made the assumption that true exposure $X_i(t)$ was equal to the observed exposure values in the cohort and multiplied $X_i(t)$ by an error term to obtain observed exposure $Z_i(t)$.

Chapter 6

Results

6.1 The effects of measurement model misspecification on simulated data

In this chapter, we will present the results obtained for each of the three main studies that were realised in this work:

1. Simulation study 1 and 2 based on the exposure data of the French cohort of uranium miners to assess the effects of measurement error on statistical inference in the context of proportional hazards models when this measurement error is not accounted for (methodology described in section 5.4)

2. The elicitation of prior information by expert knowledge for breathing rate of a miner intervening in the estimation of absorbed lung dose due to exposure to radon and its progeny (methodology described in section 5.2)

3. The estimation of the corrected risk estimate of lung cancer mortality associated with radon exposure in the French cohort of uranium miners when accounting for exposure and dose uncertainty (methodology described in 5.1 and 5.3).

6.1.1 The impact of shared and unshared measurement error on risk estimation (simulation study 1)

For the data sets considered in simulation study 1 and simulation study 2, Bayesian inference was conducted via the MCMC algorithm implemented in Python presented in section 5.3.1. After checking convergence on one data set per scenario, inference on all subsequent data sets was based on one chain with 20.000 iterations after an initial burn-in of 10.000 iterations (with a thin of one). For each of the 16 simulation models and each of the scenarios considered in simulation study 1, we estimated the average posterior median $\hat{\beta}$ for the risk coefficient β, obtained on 100 data sets. Moreover, we estimated an overall 95% credible interval ($CI_{95\%}$), which was obtained by combining the chains for the 100 data sets for each simulation model and scenario and determining the 2.5 and 97.5 quantiles of the corresponding pooled chain. Moreover, we determined the relative bias of $\hat{\beta}$, given by $\frac{(\hat{\beta}-\beta)}{\beta}$. Finally, we calculated the coverage rates of the 95% credible intervals, which were obtained by counting the proportion of the 100 replicates

Table 6.1: Average posterior median ($\hat{\beta}$), overall 95% credible intervals ($CI_{95\%}$), relative bias and coverage rate for 100 data sets generated according to the Cox model $\mathcal{D}_{S2}$ with different measurement error characteristics and a true risk coefficient of $\beta = 2$ per 100 WLM

Type of sharing	Type of error	Error variance	$\hat{\beta}$	$CI_{95\%}$	Relative bias	Coverage rate
unshared	Berkson	0.1	1.81	[1.64; 1.99]	-0.10	0.10
		0.9	1.25	[0.97; 1.49]	-0.38	0.00
	classical	0.1	1.75	[1.55; 1.93]	-0.13	0.02
		0.9	0.83	[0.45;1.21]	-0.59	0.00
between	Berkson	0.1	1.82	[1.62; 2.01]	-0.09	0.22
		0.9	1.25	[1.03; 1.47]	-0.38	0.00
	classical	0.1	1.75	[1.53; 1.94]	-0.13	0.05
		0.9	0.80	[0.44; 1.16]	-0.60	0.00
within	Berkson	0.1	1.45	[1.12, 1.69]	-0.28	0.00
		0.9	0.76	[0.54; 0.97]	-0.62	0.00
	classical	0.1	1.33	[1.04; 1.58]	-0.34	0.00
		0.9	0.39	[0.17; 0.62]	-0.81	0.00
both	Berkson	0.1	1.46	[1.11; 1.78]	-0.27	0.00
		0.9	0.77	[0.54; 1.00]	-0.62	0.00
	classical	0.1	1.42	[1.04; 1.72]	-0.29	0.00
		0.9	0.49	[0.13; 0.86]	-0.76	0.00
none	none	0	1.96	[1.80; 2.13]	-0.02	0.95

for which the 95% credible interval included the true value of the coefficient β where β is the risk coefficient which served to generate the data.

As can be seen in Table 6.1, in the Cox proportional hazards model (i.e. disease model $\mathcal{D}_2$), exposure uncertainty shared within workers created more relative bias in risk estimates and smaller coverage rates than exposure uncertainty shared between workers. For instance, concerning small measurement error of both Berkson and classical nature, the relative bias was more than twice as large when this error was shared within workers rather than between workers. In general, the impact of unshared uncertainty and uncertainty shared between workers on risk estimation was comparable. Error components, which were both shared between and within individuals, produced about as much bias as error components that were only shared within individuals.

As can be seen in Table 6.2, the relative bias introduced by measurement error when the data were generated according to an EHR model was smaller than the bias we observed when data

were generated according to the Cox model ($\mathcal{D}_{S2}$). For large measurement error in the EHR model, we observed the same pattern as for the Cox model where measurement errors shared within workers caused more relative bias and lower coverage rates than unshared measurement error or measurement error that was only shared between workers. For small measurement error, this tendency was less evident.

In general, classical measurement caused more relative bias and smaller coverage rates than Berkson error and large measurement error caused more relative bias and smaller coverage rates than small measurement error, regardless of the disease model and regardless of whether exposure uncertainty was shared or unshared.

As expected, when data were generated without exposure measurement error, the coverage rates of the 95% credible intervals were very close to 95%.

The statistical power was estimated to be 100% for all scenarios, both for data generated according to the Cox model and according to the EHR model.

Table 6.2: Average posterior median ($\hat{\beta}$), overall 95% credible intervals ($\text{CI}_{95\%}$), relative bias and coverage rate for 100 data sets generated according to the EHR model $\mathcal{D}_{S1}$ with different measurement error characteristics and a true risk coefficient of $\beta = 5$ per 100 WLM

Type of sharing	Type of error	Error variance	$\hat{\beta}$	$\text{CI}_{95\%}$	Relative bias	Coverage rate
unshared	Berkson	0.1	4.87	[3.07; 7.68]	-0.03	0.93
		0.9	4.65	[2.89; 7.18]	-0.07	0.91
	classical	0.1	4.88	[3.13;7.47]	-0.02	0.94
		0.9	4.34	[2.71; 6.70]	-0.13	0.78
between	Berkson	0.1	4.77	[3.14; 7.11]	-0.05	0.99
		0.9	4.69	[2.91; 7.31]	-0.06	0.93
	classical	0.1	4.79	[3.04; 7.35]	-0.04	0.93
		0.9	4.44	[2.82; 6.72]	-0.11	0.85
within	Berkson	0.1	4.88	[3.13; 7.47]	-0.02	0.94
		0.9	3.98	[2.43; 6.23]	-0.20	0.73
	classical	0.1	4.75	[3.01; 7.31]	-0.05	0.91
		0.9	3.03	[1.86; 4.71]	-0.39	0.13
both	Berkson	0.1	4.88	[3.11; 7.69]	-0.02	0.94
		0.9	3.86	[2.19; 6.59]	-0.23	0.55
	classical	0.1	4.76	[2.94; 7.29]	-0.05	0.92
		0.9	3.15	[1.62; 5.25]	-0.37	0.25
none	none	0	4.90	[3.14; 7.45]	-0.02	0.96

6.1.2 The effects of measurement error characteristics on the shape of the exposure-response curve (simulation study 2)

As can be seen in Figure 6.1, exposure-response curves for data generated according to the Cox model $\mathcal{D}_{S2}$ with no measurement error ($\mathcal{M}_{S0}$) or unshared and homoscedastic Berkson error ($\mathcal{M}_{S1}$) were close to linear on the log-scale. Heteroscedastic unshared error ($\mathcal{M}_{S2}$) appeared to create a slightly non-linear association on the log-scale. Finally, for data generated according to the measurement models incorporating shared sources of uncertainty ($\mathcal{M}_{S3}$ and $\mathcal{M}_{S4}$), we can observe a substantial attenuation of the exposure-response curve at high exposure values. Table 6.3 shows the risk estimates obtained on 100 data sets generated according to different disease and measurement models and estimated by the piecewise-linear disease models $\mathcal{D}_{E5}$ and $\mathcal{D}_{E6}$. It also highlights the impact of different error characteristics on the choice of the disease

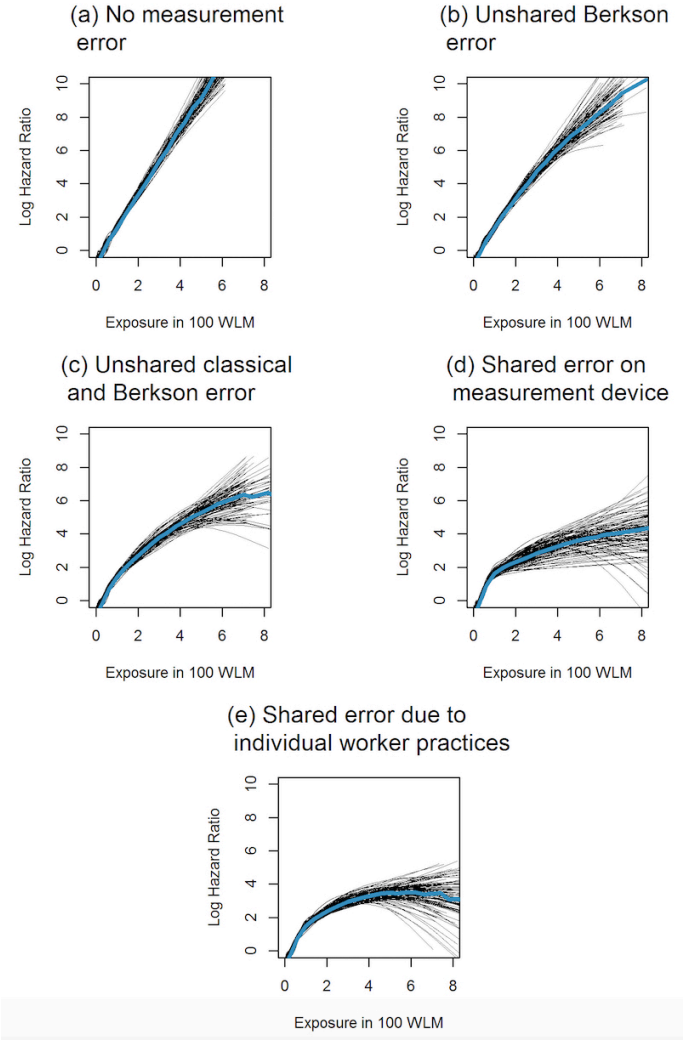

Figure 6.1: Estimated exposure-response curve when fitting the Cox model $\mathcal{D}_{E4}$ based on natural cubic splines where data are generated according to the Cox model $\mathcal{D}_{S2}$ with a risk coefficient of $\beta = 2$. (a) $\mathcal{M}_{S0}$, i.e., no measurement error (b) $\mathcal{M}_{S1}$, i.e., unshared and homoscedastic Berkson error, (c) $\mathcal{M}_{S2}$, i.e., unshared error of Berkson and classical type (d) $\mathcal{M}_{S4}$, i.e., heteroscedastic error with a shared classical component describing the imprecision of the measurement device and (e) $\mathcal{M}_{S3}$, i.e., heteroscedastic error with a shared Berkson component describing individual worker practices

Table 6.3: Comparison of risk estimates $\hat{\beta}$ (i.e. average posterior medians) with 95% credible intervals where data are generated according to different disease and measurement models. $\text{DIC}_{EHR} < \text{DIC}_{Cox}$ gives the percentage of datasets for which the Deviance Information Criterion (DIC) was smaller when fitting the linear Excess Hazard Ratio (EHR) model ($\mathcal{D}_{E1}$) when the true model was the Cox model ($\mathcal{D}_{S2}$) and vice versa for $\text{DIC}_{Cox} < \text{DIC}_{EHR}$. The difference in DIC is calculated as difference between the DIC of the EHR model and the DIC of the Cox model.

Disease model	**Model $\mathcal{M}_{S0}$** No error	**Model $\mathcal{M}_{S1}$** Unshared Berkson error	**Model $\mathcal{M}_{S2}$** Unshared heteroscedastic Berkson and classical error	**Model $\mathcal{M}_{S3}$** Heteroscedastic worker practices	**Model $\mathcal{M}_{S4}$** Heteroscedastic shared device
Data generated according to the Cox model ($\mathcal{D}_{S2}$) with $\beta = 2$					
Risk estimate $\hat{\beta}$ in the linear Cox model ($\mathcal{D}_{E2}$)	1.97 [1.78; 2.16]	1.67 [1.50; 1.87]	1.23 [1.00; 1.42]	0.77 [0.59; 0.98]	0.57 [0.21; 1.06]
Risk estimates in the piecewise-linear Cox model ($\mathcal{D}_{E6}$)					
$\hat{\beta}_1$ (under 100 WLM)	1.98 [1.57; 2.40]	2.08 [1.65; 2.49]	2.21 [1.78; 2.61]	2.33 [1.93; 2.70]	2.50 [2.06; 2.91]
$\hat{\beta}_2$ (over 100 WLM)	1.96 [1.68; 2.26]	1.49 [1.22; 1.80]	0.92 [0.64; 1.18]	0.40 [0.20; 0.63]	0.31 [0.06; 0.68]
$\text{DIC}_{EHR} < \text{DIC}_{Cox}$	0%	0%	34 %	100%	99%
Difference in DIC	-216.08	-142.13	-15.17	107.24	169.16
Data generated according to the EHR model ($\mathcal{D}_{S1}$) with $\beta = 5$					
Risk estimate $\hat{\beta}$ in the in the linear EHR model ($\mathcal{D}_{E1}$)	4.90 [3.24; 7.62]	4.71 [3.08; 7.19]	4.44 [2.93; 6.81]	4.07 [2.49; 6.28]	4.11 [2.26; 7.21]
Risk estimates in the piecewise-linear EHR model ($\mathcal{D}_{E5}$)					
$\hat{\beta}_1$ (under 100 WLM)	4.95 [2.83; 8.33]	4.81 [2.91; 7.67]	4.75 [2.79; 7.59]	4.73 [2.77; 7.64]	5.58 [3.38; 9.16]
$\hat{\beta}_2$ (over 100 WLM)	5.14 [2.06; 9.17]	4.72 [2.05; 9.21]	4.16 [1.48; 7.71]	3.09 [0.69; 6.40]	2.18 [0.27; 6.43]
$\text{DIC}_{Cox} < \text{DIC}_{EHR}$	0%	0%	0%	0%	0%
Difference in DIC	93.76	87.98	85.62	83.10	132.64

model. These results confirm that when failure times are generated according to the linear Cox model ($\mathcal{D}_2$) and the error-prone exposure data is generated according to the unshared and heteroscedastic measurement model $\mathcal{M}_{S2}$, the slope for exposure values under 100 WLM is estimated to be more than twice as large as for exposure values over 100 WLM. For the measurement models $\mathcal{M}_{S3}$ and $\mathcal{M}_{S4}$, incorporating shared sources of uncertainty due to individual worker characteristics and the precision on the measurement device in a group-level exposure assessment strategy, the slope estimates for low exposures are about six to eight times larger than the slope estimates for high exposures in this situation.

Concerning model choice, we find that, when model choice is based on the DIC, the linear EHR

model fitted the data better than the Cox model in 34% of cases when exposure data were generated following the unshared and heteroscedastic measurement model $\mathcal{M}_{S2}$, even though the true disease model was the Cox model. For measurement model $\mathcal{M}_{S3}$ and $\mathcal{M}_{S4}$, on the other hand, DIC values indicated for all replicates that the linear EHR model fitted the data better than the Cox model, although data were generated according to the Cox model. Moreover, in the three scenarios using heteroscedastic measurement models ($\mathcal{M}_{S2}, \mathcal{M}_{S3}$ and $\mathcal{M}_{S4}$), risk coefficients estimated in the Cox model based on a piecewise-linear structure $\mathcal{D}_{E6}$ were overestimated for low exposures and underestimated for high exposure values when data was generated according to the Cox model. In summary, we only observed a substantial attenuation of the exposure-response curve in the Cox model when the first exposure period was characterised by a combination of unshared and shared measurement error, which was either shared within workers or both within and between workers.

When it comes to failure times generated according to the linear EHR model ($\mathcal{D}_{S1}$), the exposure-response curves plotted in Figure 6.2, which were fitted through the EHR model based on cubic splines $\mathcal{D}_{E3}$, suggest that the different patterns of shared or unshared measurement error produce a less pronounced attenuation in exposure-response curves. The risk estimates in the piecewise linear EHR model in Table 6.3 reveals that the risk for exposures under 100 WLM is estimated to be more than twice as large as the risk estimated for exposures exceeding 100 WLM when exposure data is contaminated with classical measurement error shared both between and within workers in the first exposure period to reflect the imprecision of the measurement device ($\mathcal{M}_{S4}$). For the measurement model assuming components of Berkson error shared for several years of the same worker to account for individual worker characteristics ($\mathcal{M}_{S3}$), we also observe a higher risk estimate for exposures under 100 WLM than for exposures exceeding 100 WLM in the piecewise linear EHR model. However, in both cases, the credible intervals for the parameters in the piecewise linear model are very large and overlap. In contrast to the situation where failure times were generated according to the Cox model, DIC values always indicated the linear EHR ($\mathcal{D}_{E1}$) to be the better fitting disease model when failure times were generated according to the EHR model ($\mathcal{D}_{S1}$), regardless of the measurement model used to generate the error-prone exposure data.

6.2 Prior elicitation by expert knowledge on the average breathing rate of a French uranium miner

When specifying informative prior distributions on the unknown input parameter average breathing rate in the dose model by eliciting expert knowledge, we considered six different working conditions, l, $l \in \{1, 2, 3, 4, 5, 6\}$ depending on a dose period (until or after 1977) and a type of work (hewer, underground miner, open pit miner). As described in section 5.2., we first derived individual Dirichlet distributions $f_{e,l}$ describing the uncertainty of each expert e, $e \in \{1, 2, 3\}$ on the proportion of time a miner in working condition l spent in the three different levels of physical activity a, $a \in \{1, 2, 3\}$ (i.e. sitting, in light exercise and in heavy exercise). In a second step, we used three different approaches to derive a unique probability distribution f_l describing the proportion of time a miner in working condition l spent in the different levels of physical activity by combining the data elicited by the three experts. Finally, we based the definition

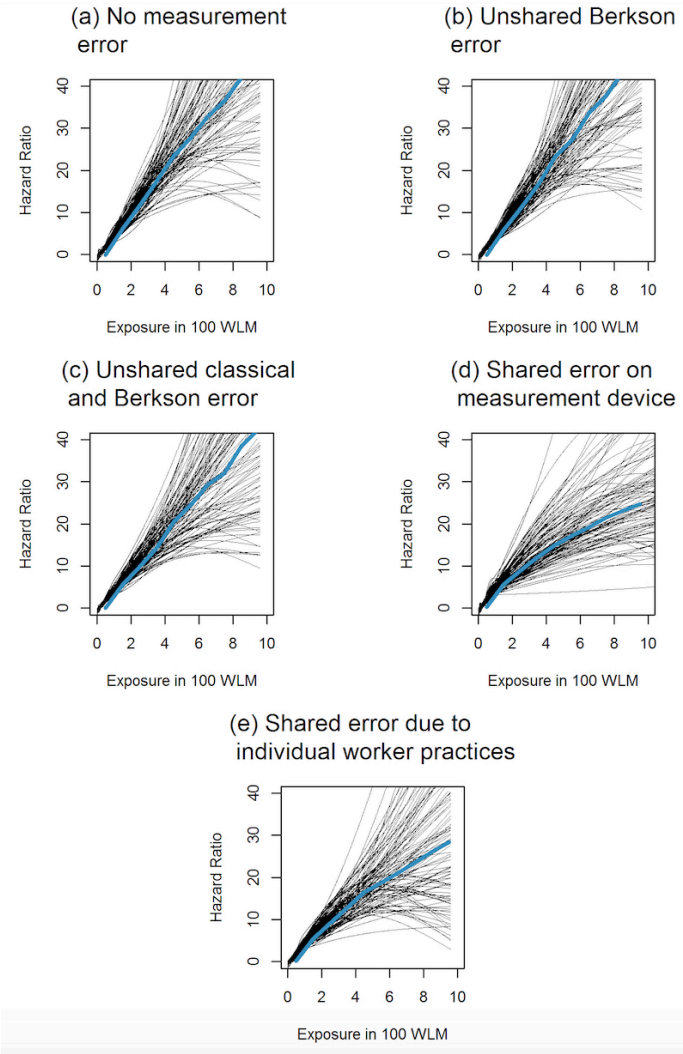

Figure 6.2: Estimated exposure-response curve when fitting the Excess Hazard Ratio (EHR) model $\mathcal{D}_{E3}$ based on natural cubic splines where data are generated according to the EHR model $\mathcal{D}_{S1}$ with a risk coefficient of $\beta = 5$. (a) $\mathcal{M}_{S0}$, i.e., no measurement error (b) $\mathcal{M}_{S1}$, i.e., unshared and homoscedastic Berkson error, (c) $\mathcal{M}_{S2}$, i.e., unshared error of Berkson and classical type (d) $\mathcal{M}_{S4}$, i.e., heteroscedastic error with a shared classical component describing the imprecision of the measurement device and (e) $\mathcal{M}_{S3}$, i.e., heteroscedastic error with a shared Berkson component describing individual worker practices

of the prior distribution on the breathing rate of a miner in the working condition, br^l, on the combined probability distribution f_l via the time activity-ventilation-approach and the values for average breathing rate in each level of physical activity recommended by the ICRP.
Expert B preferred to limit his answers to the working conditions in the French uranium mines after the mechanisation of work, as he had been employed after 1978 and he felt that he could not give information on the working conditions before that year. Similarly, Expert A chose to exclude the specification of information concerning the working condition of an open pit miner before the mechanisation, as he felt that he did not dispose of sufficient experience on this condition. Therefore, we excluded the condition of an open pit miner before the mechanisation (until 1977) and only elicited informative prior distributions on average breathing rate for the

remaining five working conditions.
In the following, we will first present the detailed steps of the derivation of an informative prior distribution on average breathing rate for the working condition "hewer after the mechanisation of work in the mines", for which all three experts had provided exploitable information. Moreover, we will present the resulting prior distributions for average breathing rate on the five working conditions for which the information of at least two experts was available. For more details on the different steps in the derivation of the informative prior distributions for all working conditions, see Appendix D.

6.2.1 Fitting independent Dirichlet distributions to describe the knowledge of each expert for a hewer after the mechanisation

Concerning the breathing rate of a hewer after the mechanisation, we disposed of 27 quantiles and 18 evaluations at the end of the elicitation procedure described in section 5.2.1. We chose to fit a Dirichlet distribution to describe the uncertainty of expert e on the proportion of time a hewer after the mechanisation spent in the different levels of physical activity via its marginal beta distributions as described in 5.2.2. The resulting marginal beta distributions are plotted in the right column of Figure 6.3. We can compare these marginal beta distributions with beta distributions which are fitted for each expert and for each level of physical activity independently. These independent beta distributions, which are shown in the left column of Figure 6.3, are fitted to the same data as the Dirichlet distributions with the only difference that they do not respect the condition that the sum of the three proportions has to be one. It is interesting to note that when considering these independent beta distributions, it seems that Expert A and Expert C have contradictory views on the time a hewer after the mechanisation spent in the different levels of physical activity. In particular concerning the proportion of time spent in light exercise, Expert A (in red) indicates that the proportion of time spent in this level of physical activity was rather small, i.e. under 25%, while Expert C (in blue) indicates that this proportion is superior to 50%. When we fit a joint Dirichlet distribution to respect the constraint that the three proportions have to sum to one for each expert, it becomes apparent that the seemingly contradictory views of Expert A and Expert C are more congruent than what could be expected from the independent beta distributions. Thereby, we observe that Expert A showed a tendency to underestimate the proportion of time spent in each activity while Expert C had a tendency to overestimate the proportion of time spent in each activity. When considering the marginal beta distributions instead of the distributions fitted on each physical activity independently, the two experts seem to give almost the same information on the proportion of time spent in heavy exercise. The fitting of a joint Dirichlet distribution results in U shaped marginal beta distributions for Expert B.

6.2.2 Combining the information on the proportion of time spent in the different levels of physical activity of the three experts for a hewer after the mechanisation

We compared three different approaches for the combination of the three individual Dirichlet

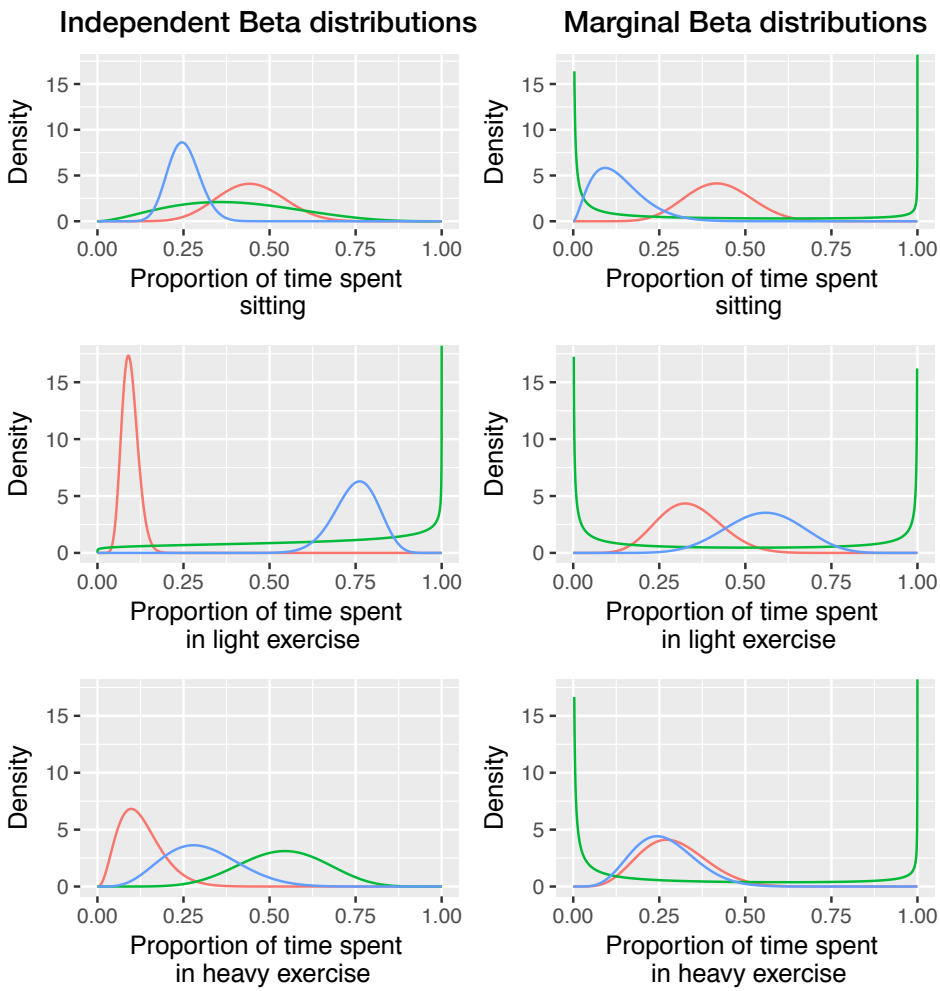

Figure 6.3: Marginal beta distributions of a common Dirichlet distribution (right column) and independent beta distribution (left column) determined via weighted least squares to describe the uncertainty of the three experts on the proportion of time a hewer after the mechanisation spent in the different levels of physical activity. The distributions fitted on the quantiles and evaluations provided by Expert A are plotted in red, for Expert B in green and for Expert C in blue.

distributions $f_{e,l}$ describing the view of each expert e, $e \in \{1, 2, 3\}$ independently, namely linear pooling, averaging and a hierarchical approach, as presented in section 5.2.3. Figure 6.4 compares the resulting marginal probability distributions combining the knowledge of all three experts on the proportion of time spent in the different levels of physical activity with the individual marginal probability distributions previously derived for each expert. As expected, linear pooling, shown in the second column of Figure 6.4, results in multimodal marginal distributions. On the other hand, averaging (third column) and the Supra-Bayesian approach for the combination of expert information (fourth column) both lead to unimodal probability distributions. By construction, the two latter approaches lead to a common probability distribution which is again a Dirichlet distribution. Concerning the results of the Supra-Bayesian approach, the specification of different rate parameters (0.01 and 0.001) for the vague exponential prior distribution on Σ_l does not seem to have a lot of influence on the results concerning a hewer

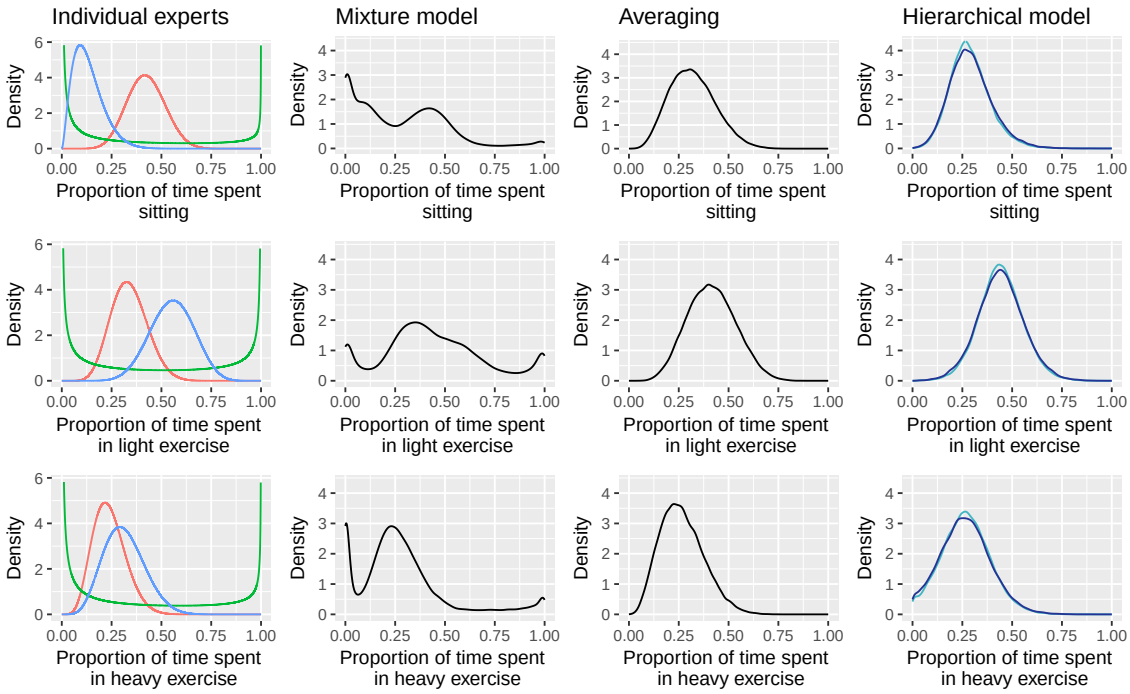

Figure 6.4: Marginal probability distributions resulting from the combination of the information provided by the three experts concerning the proportion of time spent in the different levels of physical activity via linear pooling (i.e. a mixture model), averaging and the Supra-Bayesian approach ("hierarchical model") with a rate parameter of 0.01 (dark blue) and 0.001 (light blue) for the vague exponential prior distribution on Σ_l concerning the working condition of a hewer after the mechanisation

after the mechanisation.

6.2.3 Deriving a unique prior probability distribution on the average breathing rate of a miner in the different working conditions

Finally, we used the combined joint probability distribution describing the uncertainty of the proportion of time $P_{a,l}$ a miner spent in the different levels of physical activity for a specific working condition to derive a unique prior distribution on the average breathing rate br^l of a miner in this working condition l via the time-activity-ventilation approach and the values for average breathing rate in each level of physical activity recommended by the ICRP (see section 5.2). Figure 6.5 shows the prior distributions on the average breathing rate for the five working conditions for which we disposed of information provided by at least two experts (this excludes the working condition of an open pit miner before the mechanisation for which only Expert C provided information). In general, the resulting prior distributions indicate that the average breathing rate of a miner was higher before the mechanisation than after the mechanisation. Linear pooling (in blue) leads to multimodal prior distributions on the average breathing rate for a hewer before the mechanisation, a hewer after the mechanisation and an underground miner after the mechanisation, but not for an open pit miner after the mechanisation and an underground miner before the mechanisation. For the latter two conditions, the three approaches for the combination of expert opinion result in virtually the same prior distributions on the

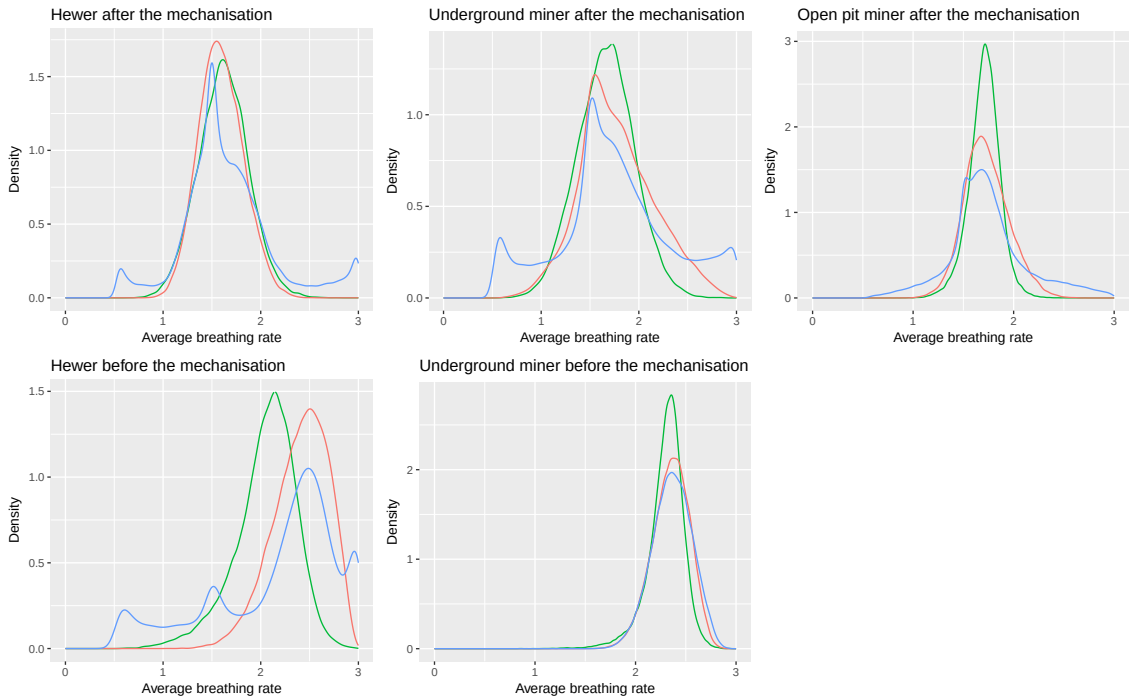

Figure 6.5: Unique prior probability distribution on the average breathing rate of a French uranium miner in the different working conditions derived via linear pooling (in blue), averaging (in red) and the Supra-Bayesian approach in green.

average breathing rate. Concerning a hewer after the mechanisation, the mode of the three alternative prior distributions are virtually the same, while the variance for the prior based on linear pooling (in blue) is larger than the variance for the prior based on the Supra-Bayesian approach (in green), which is in turn slightly larger than the variance for the prior based on averaging (in red). For the average breathing rate for a hewer before the mechanisation, the three approaches yield contrasting results. The distribution based on linear pooling (blue curve) indicates that there is a considerable probability mass on a wide range of values, in particular for values that are inferior to 1.5 m^3h^{-1}, for which the prior based on averaging (red curve) only specifies a negligible probability mass. While the blue curve and the red curve concur in their mode, the Supra-Bayesian approach can be seen to make a compromise between this mode and the non-negligible probability mass indicated by the blue curve.

It is interesting to compare the prior probability distributions on the average breathing rate of a French uranium miner that we derived through the elicitation of expert knowledge with the prior distribution based on the information available in the literature, shown in Figure 6.6. This lognormal prior distribution, that can be used for all working conditions, indicates that there is a high probability mass for values inferior to 0.54 m^3h^{-1}, which is the average breathing rate corresponding to the proportion of time a worker spent sitting. In other words, according to this prior distribution, proposed by Birchall et al. (1994) [2], it is possible that a worker had a smaller average breathing rate than if he spent his entire working day sitting. In light of the different tasks of underground miners, this seems to be rather unlikely. Moreover, this prior distribution does not reflect the intervals for the breathing rate of workers proposed by the ICRP (1.0-1.7 $\mathrm{m^3h^{-1}}$) or elaborated in the context of the alpha risk project (1.1-2.1 $\mathrm{m^3h^{-1}}$

and 0.9-1.7 m^3h^{-1} obtained by varying the time a miner spent in the different levels of physical exercise or by varying the average breathing rate for the different levels of physical activity, respectively). The prior distributions we derived from expert knowledge on the other hand, are in accordance with these intervals and respect the condition that the average breathing rate of a miner is unlikely to be inferior to 0.54 m^3h^{-1}.

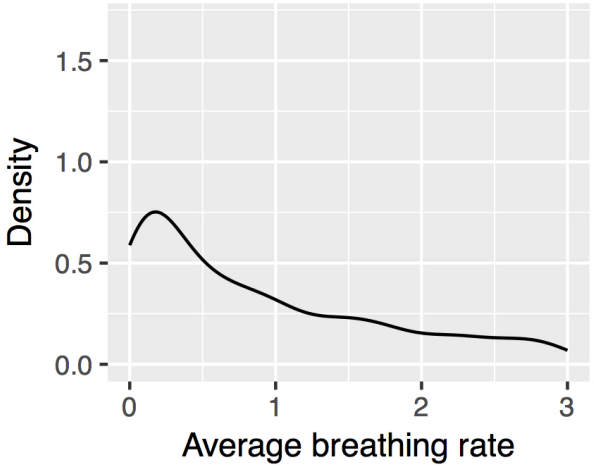

Figure 6.6: Prior distribution on the average breathing rate of a miner proposed by Birchall et al. (1994) [2]

6.3 Accounting for exposure and dose uncertainty in the French cohort of uranium miners

6.3.1 Uncorrected results on the association between radon exposure and lung cancer mortality

Comparing different disease model structures to describe the association between radon exposure and lung cancer mortality in the French cohort of uranium miners

The effects of ionising radiation on mortality are commonly modelled via the disease model $\mathcal{D}_1$, i.e. a simple linear Excess Hazard Ratio (EHR) structure. We compared this popular model structure with disease model $\mathcal{D}_2$, i.e. a log-linear proportional hazards model, which we introduced in section 5.1.1 as Cox model. As can be seen in Table 6.4, when measurement error is not accounted for in statistical inference, DIC values indicate that the linear EHR model seems to be slightly more appropriate to describe the exposure-risk relationship between radon exposure and lung cancer mortality in the French cohort of uranium miners than the linear Cox model.

Sensitivity to prior distributions on baseline hazard

As described in section 5.1.5, we either assumed informative gamma distributions based on information on lung cancer mortality in the general French population for the parameters describing

Table 6.4: Posterior means and 95% credible intervals for β and values for the Deviance Information Criterion (DIC) for the Excess Hazard Ratio (EHR) model and the Cox model. The estimates for β are given per 100 WLM.

Disease model	β	DIC
$\mathcal{D}_1$: EHR	0.88	5433.37
	[0.50; 1.36]	
$\mathcal{D}_2$: Cox model	0.27	5443.64
	[0.16; 0.37]	

Table 6.5: Median posterior values and 95% credible intervals for the post-55 and the total cohort when assuming the simple linear Excess Hazard Ratio (EHR) model $\mathcal{D}_1$ and different prior distributions for the baseline hazard parameters

Model	β_1	λ_2 in 10^{-6}	λ_3 in 10^{-6}	λ_4 in 10^{-6}
Post-55 cohort				
Informative gamma priors	1.92	1.13	5.15	8.65
	[0.93;3.22]	[0.85;1.48]	[4.29;6.13]	[6.47;11.29]
Large uniform priors	2.75	0.80	4.25	7.47
	[1.18;5.32]	[0.45;1.29]	[2.85;6.10]	[4.60;11.80]
Total cohort				
Informative gamma priors	0.88	1.26	5.24	9.84
	[0.50;1.36]	[0.98;1.58]	[4.46;6.11]	[7.92;12.13]
Large uniform priors	0.88	1.19	5.11	10.28
	[0.44;1.49]	[0.82;1.66]	[3.96;6.46]	[7.72;13.42]

the baseline hazard after 40 years of age, i.e. $\lambda_2, \lambda_3, \lambda_4$ or flat prior priors in the form of uniform distributions. Table 6.5 shows the estimates of the parameters of the disease model for the total and the post-55 cohort when assuming the simple linear EHR model without effect modification and without measurement error correction. Excess hazard estimates are given per 100 WLM. There was no impact of the prior choice for the baseline hazard parameters on risk estimates for the total cohort with only slightly narrower credible intervals for β when assuming informative gamma distributions. For the post-55 cohort, on the other hand, the definition of prior distributions for baseline hazard seemed to be quite influential with an estimated risk estimate of 1.92 assuming informative gamma distribution and a risk estimate of 2.75 when assuming uniform priors. Under the assumption of uniform priors, all baseline hazard estimates in the post-55 cohort were clearly inferior to the corresponding values estimated on the full cohort whereas they are comparable when assuming informative priors. Based on these results, we chose flat uniform distributions on the baseline hazard parameters for the post-55 cohort for all subsequent analyses. Concerning the total cohort, we based all analyses on informative gamma priors when obtaining uncorrected risk estimates and when accounting for unshared measure-

ment error. Moreover, to obtain uncorrected risk estimates and when accounting for unshared measurement error, we chose a left-truncation of the prior distribution on $\boldsymbol{\beta}$ to guarantee the positivity of the instantaneous hazard rate $h_i(t)$.

Accounting for different effect-modifying variables in the exposure-response relationship

Table 6.6: Median posterior values and 95% credible intervals for the Excess Hazard Ratio per 100 WLM for different disease models for the total cohort without measurement error correction

Model	EHR per 100 WLM	CI 95%	DIC
$\mathcal{D}_1$: Linear	0.88	[0.50;1.36]	5433.37
$\mathcal{D}_3$: Period of exposure			5422.49
≤ 1955	0.34	[0.04;0.83]	
> 1955	1.95	[1.16;2.93]	
$\mathcal{D}_4$: Time since exposure			5435.51
[5 - 15[years	1.49	[0.22;3.64]	
[15 - 25[years	1.14	[0.12;2.69]	
≥ 25 years	0.78	[0.31;1.40]	
$\mathcal{D}_5$: Piecewise linear			5428.14
< 50 WLM	1.99	[0.98;3.19]	
≥ 50 WLM	0.37	[0.02;0.98]	
$\mathcal{D}_6$: Piecewise linear			5426.56
< 100 WLM	1.46	[0.84;2.21]	
≥ 100 WLM	0.26	[0.01;0.88]	

As described in section 5.1.1, we implemented different Excess Hazard Ratio (EHR) models as a function of different effect modifying variables. Table 6.6 and Table 6.7 show the median posterior estimates and 95% credible intervals as well as DIC values for the different disease models for the total cohort and the post-55 cohort, respectively. Bayesian inference for the different models was conducted on three MCMC chains with 100.000 iterations after a burnin of 25.000 iterations and an adaptive phase in which we adapted the proposal distributions for all unknown quantities to obtain acceptance rates close to 0.4 for a single parameter and 0.2 for the vectors of latent variable values. For the total cohort, the excess hazard ratio was estimated to be 0.88 per 100 WLM when assuming the simple linear EHR model $\mathcal{D}_1$. Concerning the modifying effect of time since exposure ($\mathcal{D}_4$), the risk associated with recent radon exposure (i.e. in the last 15 years) induced an excess hazard ratio that was estimated to be twice as large as the excess hazard ratio for radon exposures received more than 25 years ago. However, according to

Table 6.7: Median posterior values and 95% credible intervals for different disease models for the post-55 cohort without measurement error correction

Model	EHR per 100 WLM	CI 95%	DIC
$\mathcal{D}_1$: Linear	2.75	[1.18;5.32]	2456.09
$\mathcal{D}_4$: Time since exposure			2457.56
[5 - 15[years	2.01	[0.10; 7.91]	
[15 - 25[years	4.82	[1.01; 11.04]	
$\geq$ 25 years	3.20	[0.77; 7.46]	
$\mathcal{D}_5$:Piecewise linear			2457.07
$<$ 50 WLM	3.20	[0.80;7.03]	
$\geq$ 50 WLM	2.82	[0.19;8.14]	
$\mathcal{D}_6$: Piecewise linear			2457.80
$<$ 100 WLM	2.90	[1.24;5.63]	
$\geq$ 100 WLM	9.28	[0.36;43.55]	

the DIC, it was not pertinent to include time since exposure as effect modifying variable in the total cohort. Considering the piecewise linear models ($\mathcal{D}_5$ and $\mathcal{D}_6$), we observed an important attenuation in the exposure-response curve for the total cohort with a risk estimate that was estimated to be five times larger for exposures under 50 or 100 WLM than for exposures over these breakpoints. At the same time, the risk estimate for exposures received after 1955 was also estimated to be five times larger than the risk estimate for exposures received until 1956 (see results on model $\mathcal{D}_3$). DIC values indicate that period of exposure is the most important effect modifying variable in the total cohort.

For the post-55 cohort, on the other hand, the excess hazard ratio based on the simple linear model was estimated to be 2.75 per 100 WLM. Overall, for this sub-cohort, DIC values indicate that none of the models including effect modifying variables describes lung cancer mortality better than the simple linear EHR model without effect modification. Moreover, in contrast to the full cohort, the piecewise linear models did not indicate an attenuation of the exposure-response relationship. On the contrary, the risk estimate for exposures over 100 WLM was estimated to be approximately three times larger than for exposures under 100 WLM. However, this result has to be interpreted with great caution, since the uncertainty associated with this estimate is very large. This uncertainty is due to the fact that there was only a very small number of miners exposed to more than 100 WLM in the post-55 cohort. Similarly, concerning effect modification by time since exposure, we do not observe the same pattern as in the total cohort. Indeed, on the post-55 cohort, the risk estimates associated with more recent exposure values is estimated to be smaller than for exposure received 15 to 25 years or more than 25 years ago. The category associated with the highest estimated risk coefficient on the post-55 cohort is $[15-25[$ years. However, the uncertainty associated with the risk parameters estimated when

assuming a disease model taking into account time since exposure was very large.
To further assess the reasons for observing evidence for effect modification of time since exposure for the full cohort, but not for the post-55 cohort, we analysed the data of the miners that were employed until 1955. This sub-cohort consists of all miners which are in the full cohort, but not in the post-55 cohort. We will call this cohort pre-55 cohort in the following. Table 6.8 shows the risk estimates for this sub-cohort. The risk estimates for the three exposure windows

Table 6.8: Median posterior values and 95% credible intervals for the Excess Hazard Ratio per 100 WLM for different disease models for the pre-55 cohort without measurement error correction

Model	EHR per 100 WLM	CI 95%	DIC
$\mathcal{D}_1$: Linear	0.50	[0.16;1.03]	2975.74
$\mathcal{D}_4$: Time since exposure			2978.48
[5 - 15[years	0.76	[0.13; 1.86]	
[15 - 25[years	0.58	[0.04; 1.71]	
$\geq$ 25 years	0.56	[0.04; 1.60]	
$\mathcal{D}_5$: Piecewise linear			5429.14
<50 WLM	1.47	[0.21;3.54]	
$\geq$50 WLM	0.35	[0.03;0.97]	
$\mathcal{D}_6$: Piecewise linear			5426.56
<100 WLM	1.04	[0.30;2.21]	
$\geq$100 WLM	0.28	[0.02;0.94]	

are very similar. Similarly to the post-55 cohort, we therefore do not observe evidence for the effect modification by time since exposure in this sub-cohort either. In summary, neither the mortality data on miners employed until 1955 nor the data on miners employed after 1955 suggest that there is an effect modification of the association between radon exposure and lung cancer mortality by time since exposure. However, when considering these two sub-cohorts together in the full French cohort of uranium miners, we observe evidence for the effect modification by time since exposure. This phenomenon can be explained by the fact that the risk estimates estimated on the data of the post-55 cohort are estimated to be higher then in the pre-55 cohort and the exposures received in the post-55 cohort are more recent than the exposure received in the pre-55 cohort.

Checking the proportional hazards assumption

It was merely possible to test the proportional hazards assumption in the Cox model without the effect modification, since this model was the only one which could be estimated with the library survival in R. Based on the Schoenfeld residuals, plotted in Figure 6.7, and the Harrell test ($p = 0.16$), we did not reject the hypothesis of proportional hazards. Therefore, there is

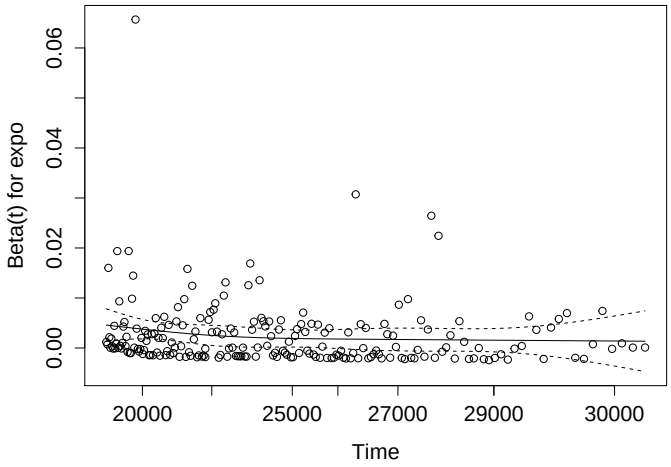

Figure 6.7: Schoenfeld residuals for the Cox model $\mathcal{D}_2$

no evidence for the time-dependence of the regression coefficient β associated with cumulated radon exposure in this model. This result suggests that the proportional hazards assumption might as well be fulfilled for the disease model which uses an EHR structure. Even if the two models differ in the exact structure of the exposure-risk relation, we believe that the parameter β should be either time-varying in both or in neither of them.

6.3.2 Accounting for unshared exposure uncertainty

In a first attempt to account for exposure uncertainty in the French cohort of uranium miners, we combined the different disease models presenting an EHR structure with measurement model $\mathcal{M}_1$ described in section 5.1.2 and the exposure model described in section 5.1.3. The resulting hierarchical model presented in the Directed Acyclic Graph (DAG) in Figure 5.6 allows to obtain risk estimates for the total and the post-55 cohort that are corrected for unshared Berkson and classical measurement error according to period of exposure. Table 6.9 and 6.10 show the resulting posterior medians and 95% credible intervals for the different disease models. When comparing these results with the naive estimates obtained when not accounting for measurement error in Table 6.6 and Table 6.7, we do not observe substantial differences in risk estimates and in credible intervals. In particular, concerning the full cohort, we continue to observe a substantial effect modification by the period of exposure and there is evidence for the non-linearity of the exposure-response relationship when testing the piecewise linear models. When comparing the estimates in the linear EHR model ($\mathcal{D}_1$), we observe a small increase in the linear excess hazard ratio when accounting for unshared exposure uncertainty in the full cohort.

Concerning the post-55 cohort, accounting for unshared measurement error did not change the posterior median of β in the linear EHR model ($\mathcal{D}_1$), but only slightly increased the boundaries of the 95% credible interval. Similarly, in the other disease models accounting for different effect modifying variables, we did not observe a marked difference in risk estimates when accounting for measurement error. The definition of less informative prior distributions for the exposure model did not substantially change risk estimates in the post-55 and the full cohort.

Table 6.9: Median posterior values and 95% credible intervals for the Excess Hazard Ratio per 100 WLM for different disease models when accounting for unshared measurement error in the total cohort

Model	EHR per 100 WLM	CI 95%	DIC
$\mathcal{D}_1$: Linear	0.90	[0.51;1.41]	5433.30
$\mathcal{D}_3$: Period of exposure			5423.59
until 1956	0.31	[0.02;0.79]	
after 1955	2.06	[1.34; 3.00]	
$\mathcal{D}_4$: Time since exposure			5435.31
[5 - 15[years	1.88	[0.27; 4.59]	
[15 - 25[years	1.35	[0.17; 3.41]	
$\geq$ 25 years	0.74	[0.27; 1.57]	
$\mathcal{D}_5$: Piecewise linear			5424.38
$<$ 50 WLM	2.12	[1.07;3.37]	
$\geq$ 50 WLM	0.34	[0.02;0.98]	
$\mathcal{D}_6$: Piecewise linear			5422.50
$<$ 100 WLM	1.57	[0.92;2.38]	
$\geq$ 100 WLM	0.23	[0.01;0.84]	

6.3.3 Accounting for shared exposure uncertainty

In a second step, we assumed the more complex measurement model $\mathcal{M}_2$ to account for Berkson error shared within workers to reflect errors due to individual job conditions and worker practices in the first two exposure periods which were characterised by a group-level exposure assessment.

Sensitivity to prior distributions on baseline hazard parameters λ_2, λ_3, λ_4

We did an additional sensitivity analysis with respect to the prior distribution when accounting for shared Berkson error in the full cohort. Table 6.11 shows the estimates of the parameters of the disease model for the total cohort when assuming the simple linear EHR model without effect modification when accounting for Berkson error shared within workers for the first and the second exposure period. Contrary to the results obtained when this shared error component was not accounted for, we can observe a small impact of the prior choice for the baseline hazard parameters on risk estimates for the total cohort. Although this influence is by far less pronounced than for the effect observed for the post-55 cohort when measurement error is not accounted for (see section 6.3.1), it suggests that the prior on the baseline hazard parameters can have an influence on risk estimation in the full cohort. In line with the hypothesis of a healthy worker effect, we would expect a lower baseline mortality rate in an occupational cohort than in

Table 6.10: Median posterior values and 95% credible intervals for the Excess Hazard Ratio per 100 WLM for different disease models when accounting for unshared measurement error in the post-55 cohort

Model	EHR per 100 WLM	CI 95%	DIC
$\mathcal{D}_1$: Linear	2.75	[1.20;5.36]	2456.09
$\mathcal{D}_4$: Time since exposure			2457.55
[5 - 15[years	2.03	[0.09;7.86]	
[15 - 25[years	4.82	[1.01;10.89]	
$\geq$ 25 years	3.19	[0.79;7.49]	
$\mathcal{D}_5$: Piecewise linear			2456.85
$<$ 50 WLM	3.22	[0.82;7.11]	
$\geq$ 50 WLM	2.72	[0.16;7.98]	
$\mathcal{D}_6$: Piecewise linear			2458.09
$<$ 100 WLM	2.85	[1.20;5.60]	
$\geq$ 100 WLM	9.50	[0.36;43.66]	

Table 6.11: Posterior medians and 95% credible intervals of the Excess Hazard Ratio (EHR) per 100 WLM and of the baseline hazard parameters λ_2, λ_3 and λ_4 obtained for the total cohort when assuming a hierarchical model based on the combination of the linear EHR disease model $\mathcal{D}_1$ and the measurement model $\mathcal{M}_2$ and different prior distributions for the baseline hazard parameters λ_2, λ_3 and λ_4

Prior	EHR per 100 WLM	λ_2 in 10^{-6}	λ_3 in 10^{-6}	λ_4 in 10^{-6}
Informative gamma priors	0.97	1.25	5.20	9.78
	[0.54;1.52]	[0.97;1.57]	[4.41;6.07]	[7.86;12.04]
Large uniform priors	0.99	1.17	5.00	10.12
	[0.48;1.73]	[0.80;1.63]	[3.85;6.36]	[7.55;13.25]

the general population. The fact that we observe a lower baseline mortality rate in the post-55 cohort but not in the full cohort could either be explained by systematic differences in these two cohorts (which are rather unlikely) or by the fact that there is an excess in lung cancer mortality in the full cohort that is caused by exposure to radon, but which cannot be explained by the observed radon exposure values. Indeed, due to the uncertainty in the exposure reconstruction before 1956, which was far greater and more complex than after 1956, there could be a substantial underestimation of the risk of lung cancer mortality before 1956 (as suggested by simulation study 2). If we try to correct for this underestimation, it seems more suitable to choose flat prior distributions on the baseline hazard parameter instead of using the information on the

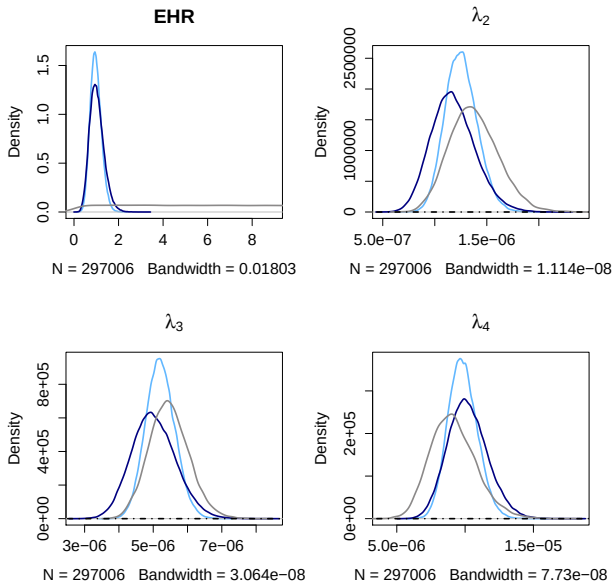

Figure 6.8: Posterior distributions for selected parameters of the disease model under flat (dark blue) and informative (light blue) prior distributions and their corresponding prior distributions (dot-dashed line and grey line, respectively) when accounting for Berkson error shared within miners for the first two exposure periods in the full cohort.

general French male population. Figure 6.8 illustrates this association between the estimated baseline hazard parameters and the risk estimate. Based on these arguments, we chose flat uniform distributions on the baseline hazard parameters for the total cohort for all subsequent analyses. Contrary to the analyses we conducted to obtain uncorrected risk estimates and when accounting for unshared measurement error, we chose no left truncation of the prior distribution on $\boldsymbol{\beta}$. Instead, we used a more generic approach to assure the positivity of the baseline hazard rate $h_i(t)$ for all miners i by setting the likelihood equal to 0 whenever there was one or more miner for which this condition was not respected when calculating the acceptance ratios used in the Metropolis-Hastings updates. Based on these changes in the definition of the prior distributions for the full cohort, we will present in the following a comparison of the uncorrected risk estimates obtained for the full cohort and for the risk estimates when accounting for Berkson error shared within miners.

Comparison of uncorrected risk estimates and when accounting for shared Berkson error

Table 6.12 shows the comparison of the uncorrected risk estimates and the risk estimates corrected for Berkson error shared within workers for the full cohort. Concerning the simple linear EHR model $\mathcal{D}_1$, we observed a marked increase in the risk estimate for lung cancer mortality and a widening of the 95% credible intervals. However, the effect modification by period of exposure and the attenuation of the exposure-response relationship seem to persist when we account for shared Berkson error in the full cohort. We can observe an increase in the risk estimates for exposure received until 1955 and after 1955 in disease model $\mathcal{D}_3$, but risk associated with the latter is still estimated to be more than five times greater than the risk associated with the

Table 6.12: Posterior medians and 95% credible intervals for different disease models for the total cohort. Comparison of uncorrected risk estimates and risk estimates accounting for Berkson error shared within miners for the two first exposure periods (measurement model $\mathcal{M}_2$).

Model	Uncorrected		Corrected	
	EHR per 100 WLM	DIC	EHR per 100 WLM	DIC
$\mathcal{D}_1$: Linear		5435.08		5369.30
	0.87 [0.42;1.50]		0.99 [0.48;1.73]	
$\mathcal{D}_3$: Period of exposure		5428.12		5346.11
until 1955	0.31 [-0.01;0.80]		0.39 [0.02;1.02]	
after 1955	1.94 [1.15;2.96]		2.21 [1.20;3.58]	
$\mathcal{D}_5$: Piecewise linear		5429.31		5368.40
< 50 WLM	2.57 [1.18;4.57]		2.07 [1.10;3.54]	
≥ 50 WLM	0.30 [-0.15;0.98]		0.25 [-0.13;1.11]	

former. Concerning the risk estimates in the piecewise linear EHR model $\mathcal{D}_5$, we observe both a decrease in the slope estimated for low exposures and high exposures and considerably more narrow credible intervals for the slope estimated for low exposures.
As can be seen in Table 6.13, we do not observe a notable modification in risk estimation when accounting for shared Berkson error in the post-55 cohort.

Table 6.13: Posterior medians and 95% credible intervals for the linear Excess Hazard Ratio (EHR) model for the post-55 cohort. Comparison of uncorrected risk estimates and risk estimates accounting for Berkson error shared within miners for the two first exposure periods (measurement model $\mathcal{M}_2$).

Model	Uncorrected		Corrected	
	EHR per 100 WLM	DIC	EHR per 100 WLM	DIC
$\mathcal{D}_1$: Linear		2455.93		2418.72
	2.69 [1.16;5.31]		2.68 [1.12;5.31]	

6.3.4 Accounting for exposure and dose uncertainty in the post-55 cohort

Table 6.14 shows posterior medians and 95% credible intervals of the Excess Hazard Ratio of death by lung cancer in Gray for the post-55 cohort when accounting neither for dose uncertainty nor measurement error (on the left), when accounting for dose uncertainty (in the middle) and when accounting for both exposure and dose uncertainty (on the right). We see a slight decrease

in the EHR when accounting for dose uncertainty via the dose model (A1) defined in section (5.1.4) but not for exposure uncertainty. When simultaneously accounting for dose and exposure uncertainty due to Berkson error shared within workers (measurement model $\mathcal{M}_2$), this increase is less pronounced and we observe a narrowing of the 95% credible intervals. The corresponding hierarchical model is illustrated in the DAG in Figure 5.9 in section 5.1.4.

Table 6.14: Posterior medians and 95% credible intervals for the total cohort. The values of the Deviance Information Criterion (DIC) are given for the model accounting for Berkson error shared within miners for the two first exposure periods (measurement model $\mathcal{M}_2$).

Model	Uncorrected		Dose uncertainty		Exposure and dose uncertainty	
	EHR per Gray	DIC	EHR per Gray	DIC	EHR per Gray	DIC
$\mathcal{D}_1$: Linear	5.92 [2.48;11.60]	2456.35	5.60 [2.30;11.30]	2418.72	5.70 [2.40;11.10]	2415.96

Chapter 7

Discussion

7.1 Summary and short-term perspectives

The focus of this work was on the treatment of measurement error in an occupational cohort study in radiation epidemiology. The combination of the two terms "cohort" and "occupational" may lead to complex structures of exposure measurement error. Indeed, in a given occupational context, we are often faced with a variety of exposure conditions and a multitude of methods of exposure assessment, which are likely to induce complex measurement error structures. The temporal dimension of a cohort study can further complicate this situation. When studying the association between the exposure to a given chemical or physical agent and a health outcome, it is common to consider cumulated exposure as the most suitable exposure metric, leading to an explanatory variable that is inherently time-dependent. While it is already difficult to adequately model the time-dependence of cumulated exposure in an epidemiological study, it is even more challenging to account for measurement error in this time-dependent context. The modelling of measurement error in an occupational cohort study may therefore involve various challenges. These challenges include heteroscedastic errors, the combination of Berkson and classical measurement error components, the possibility of measurement error to be shared between workers, within workers or both and the modelling of measurement error occurring in a time-dependent variable.

7.1.1 The effects of exposure measurement error on statistical inference

In light of the challenges that are involved in the modelling of the error structures that arise in occupational cohort studies, we can ask ourselves whether it is really worthwhile to account for measurement error in these studies. To answer this question, we assessed the effects of different measurement error structures on risk estimation and on the observed exposure-response relationship in an occupational cohort when measurement error is not accounted for in statistical inference. More precisely, we conducted two simulation studies based on the exposure data of the French cohort of uranium miners. The aim of the first simulation study was to compare the impact of shared and unshared measurement error on risk estimation in proportional hazards models. The aim of the second simulation study was to analyse the effects of complex error

structures, which are likely to arise in an occupational cohort, on the shape of the exposure-response curve. Contrary to many other authors, who addressed the problem of measurement error in occupational cohort studies, we treated cumulative exposure as a time-varying variable in a proportional hazards model in these simulation studies instead of considering the total sum of exposure values received by a worker during his entire career to be known at study entry. While the generation of failure times as a function of a time-varying exposure variable required the implementation of a sophisticated method for data generation based on piecewise-exponential variables, it allowed us to disentangle the effects of different components of exposure uncertainty. In particular, the treatment of cumulative exposure as a time-varying variable allowed us to compare uncertainty components that were unshared, shared among several miners, shared for several exposure values received by the same miner or both in. Contrary to common claims that error components shared between individuals cause more impact in risk estimation than unshared error components [119, 138, 427], we found that these two types of error components resulted in comparable relative bias and coverage rates for risk estimation in proportional hazards models. In contrast, we found that measurement error components shared for several years of the same worker had more impact on statistical inference than unshared measurement error components or measurement error that is shared between workers. In accordance with previous findings on the relative importance of measurement error in linear and log-linear models [433], we found that the impact of measurement error was more important in the Cox-model than in the EHR model. We observed relative biases on risk estimates that were consistent with the results obtained by Bender et al. (2005) [31] and Küchenhoff et al. (2007) [32], who studied the impact of unshared additive and multiplicative Berkson error for the Cox model on exposure data of the German cohort of uranium miners. However, our findings contrast with those obtained by Allodji et al. (2012) [34, 35]. The authors of the latter study found that large unshared Berkson error and small unshared classical measurement error could introduce large biases in risk estimation when conducting frequentist inference in Poisson regression based on an Excess Relative Risk structure on the exposure data of the French cohort of uranium miners. To better understand this discrepancy, we conducted a third simulation study to compare different methods of the generation of Berkson error, which is described in the paper "A cautionary comment on the generation of Berkson error in epidemiological studies" submitted to Radiation & Environmental Biophysics (see Appendix F). For the generation of Berkson error, we propose an approach that is different from the approach chosen by Allodji et al. (2012) [34, 35]. In this third simulation study, we show that the exposure data generated by our method fulfil the properties of Berkson error, contrary to the method used by Allodji et al. (2012) [34, 35], which inadvertently produces biased classical measurement error. This finding is confirmed theoretically. These results suggest that the discrepancy between our results and the results obtained by Allodji et al. (2012) [34, 35] could be explained by an error in their method for the generation of Berkson error. Finally, contrary to a widespread belief that Berkson error does not bias parameter estimates [3, 175, 147, 119], we found that components of Berkson error can lead to large biases in risk estimation when they are shared within workers.

In accordance with the results obtained by Steenland et al. (2015) [14], we only observed a mild attenuation of the exposure-response curve in the Cox model when assuming a structure of unshared error in which the magnitude of error and the magnitude of exposure were greatest

for the earliest years of exposure. However, we observed a marked attenuation of the exposure-response relationship for high exposure values when we assumed shared error components due to the imprecision of the measurement device or due to individual worker practices in the exposure periods that were characterised by a method of group-level exposure assessment. This finding casts doubt on the common practice of modelling measurement error on the sum of the exposure values received in the entire working career of a worker instead of modelling them on the daily, monthly or annual exposure values. Indeed, under this simplifying assumption, it is impossible to distinguish shared and unshared measurement error components in the exposure history of a worker. Finally, we found that distortions in the exposure-response relationship were more severe when data were generated according to the Cox model, rather than according to the EHR model. Moreover, when failure times were generated according to the Cox model and observed exposure values were contaminated with shared and unshared error, DIC values identified the EHR model as the model that best fitted the data. On the one hand, the robustness to measurement error makes the EHR model, which is often considered the best model to describe the effects of ionising radiation on mortality, attractive for risk modelling in epidemiological studies. On the other hand, this finding casts doubt on the possibility to identify a "true disease model" to describe the exposure-risk relationship when risk estimates are not corrected for all sources of exposure uncertainty.

In summary, the results of these simulation studies underline the importance of both a careful characterisation of all components of exposure measurement error in occupational studies and the realistic modelling of the identified error structures, in particular in cases where there are exposure periods that were characterised by a retrospective exposure reconstruction.

7.1.2 The use of a hierarchical approach to describe exposure measurement error

If, based on the results we observed on the two simulation studies we conducted, we can conclude that it is important to account for the complex patterns of exposure uncertainty arising in an occupational cohort, the next question we can ask ourselves is: How can the modelling of the complex error structures arising in an occupational cohort study be achieved? The answer we promote in this work is the use of a hierarchical approach. This hierarchical approach allows the combination of sub-models where different sources of uncertainty, including exposure and dose uncertainty, are accounted for at different levels of the hierarchy. At the cost of conditional independence assumptions, hierarchical models allow for the combination of simple sub-models into a unique and coherent framework to model complex phenomena, and, at the same time, for a clear interpretation of the parameters involved in the different sub-models. The most important strength of the hierarchical approach to account for exposure and dose uncertainty is arguably its flexibility in the modelling of complex structures of measurement error. It is broadly agreed upon that the adoption of simplified measurement models may cause erroneous inferences. Stefanski et al. (2002) [434] state that assuming an incorrect measurement model may potentially cause problems as great as those created by ignoring measurement error. Carroll (2012) [8] even goes so far as to suggest that for a statistician, the erroneous inferences which may be caused by an incorrect measurement model are worse than going to jail. Despite these dangers, many studies that choose classical methods for measurement error correction

only account for simplified measurement error structures and do not fully account for all components of exposure uncertainty. For instance, many studies account for measurement error in the exposure of interest, even though there are error-prone confounding variables. While it is difficult to account for measurement error in several variables, this choice may lead to residual confounding, because the effects of the error-prone confounding variables are only partially adjusted for. Moreover, it is common to assume homoscedastic measurement errors, although there is evidence of heteroscedasticity [435] or to assume non-differential measurement errors in case-control studies even though recall bias is likely to lead to differential measurement error. Finally, in the context of occupational cohort studies, it is common to assume that the exposure received by a worker during his entire working career is known at study entry [31, 32, 14] and to consider that all components of exposure uncertainty can be described by unshared measurement error [34, 35]. In contrast to classical methods for the correction of measurement error like simulation extrapolation (SIMEX) or regression calibration, the hierarchical approach provides a flexible framework for the modelling of complex structures of measurement error. In particular, it is straightforward to account for measurement error structures including systematic bias, the treatment of multiple mismeasured covariates [191] and the modelling of shared and unshared exposure and dose uncertainty in a time-varying variable in a coherent framework. Moreover, the hierarchical model can be further extended to account for uncertainty in additional parameters intervening in dose calculation or for missing values in the explanatory variables, regardless of whether these values are missing at random or not. However, the flexibility of this hierarchical approach to account for measurement error correction comes at the price of additional model assumptions. As a so-called structural method for the correction of measurement error, the hierarchical approach we presented here requires the specification of an exposure model, which describes the probability distribution of the true and unknown exposure $X_i(t)$ of a miner i at time t when accounting for a component of unshared classical measurement error. Several authors have proposed flexible exposure models based on mixture models in this situation [436, 373] and in future analyses it could be interesting to study the impact of alternative exposure model specifications when accounting for exposure uncertainty in the French cohort of uranium miners. In survival analysis, the use of a structural method for the correction of measurement error additionally requires the specification of the baseline hazard function. In this work, we assumed a flexible semi-parametric model that assumes that this baseline hazard is piecewise constant with intervals that were chosen in accordance with the lung cancer mortality observed in the general French male population. We tested the influence of this modelling choice under the Cox proportional hazards model and found that there was virtually no difference in risk estimates when basing risk estimation on the partial likelihood or on the full likelihood for the full and the post-55 cohort (see the results presented in Appendix C).

7.1.3 Conducting Bayesian inference to obtain corrected risk estimates through a hierarchical model

The Bayesian approach provides a flexible and coherent framework to conduct statistical inference for hierarchical models. Moreover, in spite of the fact that the choice of a prior distribution can be perceived as encumbering, the possibility to integrate external knowledge through the

specification of informative prior distributions can have substantial advantages when conducting statistical inference for hierarchical models. In complex hierarchical models, there may be parameters in some of the sub-models that are only poorly informed by the data. In these cases, the specification of informative prior distributions is indispensable to be able to conduct statistical inference for these parameters. In order to derive informative prior distributions, we developed an elicitation task and we compared different methods for the combination of expert opinion. Another important reason for the choice of a Bayesian approach is the versatility of Markov Chain Monte Carlo (MCMC) methods. MCMC methods allow to obtain parameter estimates in almost arbitrarily complex models, under the condition that it is possible to evaluate the likelihood for a given set of parameter values. However, it has to be acknowledged that the use of MCMC methods may be time-consuming, both in the estimation process and in the development of a problem-tailored implementation in cases where it is not feasible to rely on software packages like WinBUGS, JAGS or Stan for the fitting of the chosen hierarchical model. Moreover, when conducting Bayesian inference via MCMC methods, particular attention has to be paid to the convergence of the Markov chains to the posterior distribution of interest. The random walk Metropolis-Hastings algorithm that we chose for parameter estimation can suffer from an important number of deficiencies, in particular when the aim is to sample from a high-dimensional posterior distribution. Indeed, due to its random walk behaviour, the generic Metropolis-Hastings algorithm is quite inefficient when exploring high-dimensional posterior distributions. Thereby, its use in these situations entails problems in terms of convergence to the stationary distribution and may result in very time-consuming MCMC implementations [415, 191]. Promising alternatives in this situation include Integrated Nested Laplace Approximation [415, 191], as discussed in section 5.3.1, and a MCMC implementation that relies on Hamiltonian dynamics to suppress the random walk behaviour of the generic Metropolis-Hastings algorithm through the use of first-order gradient information [437]. While we were able to achieve important gains in computational efficiency through the use of sparse matrices in the implementation of the MCMC algorithm used to conduct Bayesian inference in this work, the integration of Hamiltonian dynamics seems a promising direction to further improve this implementation in future analyses. Owing to the object-oriented implementation of the algorithm, it appears that the integration of Hamiltonian dynamics would be relatively straightforward.

While the hierarchical approach is very flexible, it is important to stay as parsimonious as possible when using this approach. Based on the results of the simulation studies that we conducted, we may venture the conclusion that we can neglect small unshared classical measurement error components (as in the case of the error arising after 1983) and unshared Berkson error components. When accounting for exposure and dose uncertainty in the post-55 cohort, we followed this strategy by only accounting for Berkson error shared within miners before 1983 and for components of dose uncertainty that were shared for several exposure values received by the same worker. While this choice led to a substantial reduction in the dimensionality of the full posterior distribution, it is questionable whether the achieved reduction was sufficient to allow for the full exploration of the associated posterior distribution. If this is not the case, the resulting corrected risk estimates can be considered as an approximate estimation of the parameters of interest at best. Additionally, there seems to be very little information in the ob-

served mortality data concerning the values of the input parameters involved in dose calculation which are specific for each miner. Based on this lack of information, it may be more adequate to make the simplifying assumption that a common value, for instance, for breathing rate can be assumed for all miners in a given job condition, in a given mining location and in a given period of exposure (i.e. before or after the mechanisation of work in the mines). Thereby, we neglect all components of dose uncertainty that are merely shared within miners and focus on the components of dose uncertainty that are both shared between and within miners. The focus on these components of dose uncertainty could thereby allow to account for systematic changes that may have occurred in the relation between true exposure $X_i(t)$ and true dose $D_i(t)$ due to the changes in the tasks that were performed in French uranium mines. At the same time, it leads to a considerable reduction in the dimensionality of the full posterior distribution and may therefore allow for the exploration of this distribution via the MCMC algorithm we implemented in this work. A welcome side-effect of a more parsimonious model structure may thereby be that it alleviates some of the difficulties of a classical MCMC implementation when conducting Bayesian inference for very high dimensional posterior distributions.

Another promising perspective is the implementation of alternative criteria for model comparison and model checking. Currently, we solely base the comparison of alternative models on the Deviance Information Criterion (DIC). While this criterion is commonly used for model comparison, it is highly criticised. The points of criticism include the fact that it is not robust to parameter transformation, its lack of consistency (that is, when the sample size grows to infinity it is not certain to select the "true model" that generated the observed data) and its weak theoretical justification [426]. As the aim of this work was to improve the risk estimates associated with lung cancer mortality in the French cohort of uranium miners by accounting for exposure and dose uncertainty and not the identification of a "true model" to describe this context, the simple use of the DIC seems to be justifiable to some extent. Indeed, the model comparisons we performed in this work were not of a primary importance and our question of interest was rather to what extent the parameter estimates in the alternative disease models changed when accounting for the different sources of uncertainty. Nonetheless, a promising perspective of this work could be the implementation of more suitable criteria that would allow for the comparison of alternative disease models. In this context, the calculation of criteria that are based on posterior model probabilities seems to be very challenging as the estimation of these posterior model probabilities is very sensitive to the choice of prior distributions and requires the integration over the values of all unknown quantities. Another option may be to assess the fit of the alternative models through the posterior predictive distribution. This posterior predictive distribution describes the probability of outcomes Y^{rep} that could have been observed given the assumed probability model and given the posterior distribution of the model parameters [332]. In the context of posterior predictive checks, hypothetical outcomes Y^{rep} can be generated and compared with the observed outcomes through a chosen test statistic informing us about the fit of the alternative models to the data. In the analysis of censored failure times, the main problem in the implementation of posterior predictive checks lies in the fact that the outcome for individual i, Y_i, is the minimum of the failure time T_i concerning the event of interest (in our case death by lung cancer) and a censoring variable C_i. As described in section 5.1.1, it is common to neglect the distribution of the censoring variable C_i with the justification that the

censoring variable does not depend on the values of the parameters of interest. However, to be able to generate hypothetical censored failure times Y_i^{rep}, $i \in \{1, \ldots, n\}$, we need to be able to generate values for the censoring variable to be able to determine Y_i^{rep} as a function of T_i^{rep}. It is unlikely to find an adequate probability model for this censoring variable and the choice of an arbitrary probability distribution for C_i, for instance an exponential distribution, is likely to have a strong impact on the values of the calculated test statistics. In this situation, we could restrict ourselves to the generation of uncensored survival times and generate these uncensored predictive values only for miners who died of lung cancer, i.e. for all miners for which $Y_i = T_i$. However, in the analysis of the mortality of the French cohort of uranium miners, an additional difficulty relies in the fact that the outcome Y_i is not only censored, but also both left- and right-truncated (as it only takes values inferior to 85 years), while T_i is only left-truncated. For a generated value T_i^{rep} that is superior to 85 years, it does not seem reasonable to set $Y_i^{rep} = T_i^{rep}$, because it violates the right-truncation of the outcome. In this case, we can either set $Y_i^{rep} = 85$ years or generate new values for T_i^{rep} until it takes a value inferior to 85 years. Both options are not entirely satisfactory, where the main problem of the latter option is that the rejection of all values of T_i^{rep} that are superior to 85 years may be very time-consuming, in particular in cases where there is only a small probability mass of the posterior predictive density for T_i that is inferior to 85 years. In light of these difficulties, a first step in the improvement of the comparison of the alternative disease models may be achieved by the Watanabe-Akaike information criterion (WAIC) [332]. While the estimation of the WAIC is essentially as straightforward as the estimation of the DIC, this criterion can be considered as an improvement of the DIC as it overcomes some of the deficiencies of the DIC, including its sensitivity to reparametrization and its lack of consistency [426].

7.1.4 Accounting for exposure and dose uncertainty in the French cohort of uranium miners

In the analysis of the lung cancer mortality associated with exposure to radon and its progeny in the French cohort of uranium miners, we used a Bayesian hierarchical approach to explicitly account for exposure and dose uncertainty. We assumed different disease and measurement models to describe the association between true radon exposure and lung cancer mortality and the association between true radon exposure and observed radon exposure, respectively.
Regarding the disease model, we found that a simple linear Excess Hazard Ratio (EHR) structure, which is classically used to model the health effects of radiation exposure, seemed to be more appropriate than the log-linear Cox model to describe the association between radon exposure and lung cancer mortality in the cohort. In a first step, we compared the fit of different models that accounted for the effect modification by other variables, but not for exposure or dose uncertainty. Similarly to Vacquier et al. (2009) [70], we found period of exposure to be the most important effect-modifying variable in the full cohort with a risk that was estimated to be about five times greater for exposures received after 1955 than for exposure received until 1955. The most plausible explanation for this finding is the difference in the quality of exposure assessment with a retrospective exposure reconstruction until 1955 and a prospective exposure

assessment of good quality after 1955. Moreover, the resulting association of high exposure rate and low-quality exposure assessment seems to create an attenuation of the exposure-response relationship for high cumulated exposure values, which is a phenomenon observed in many occupational studies [189, 438, 17, 14]. While we observed this flattening of the exposure-response curve for the total cohort, there was no evidence for a comparable attenuation for high cumulated exposure values when restricting analyses to the post-55 cohort, i.e. miners that were employed after 1955 and for which radon exposure had only been assessed in a prospective fashion. The risk associated with radon exposure estimated in this sub-cohort was substantially higher than the risk estimated in the full cohort with an uncorrected risk coefficient of 2.75 [1.18;5.32] compared to an uncorrected risk estimate of 0.88 [0.50;1.36] for the full cohort. Finally, we found evidence for a modifying effect of the time since exposure in the full cohort, but neither in the post-55 cohort nor in the sub-cohort of miners employed before 1955. If there was a genuine effect modification by time since exposure, we should be able to observe it at least for one of the two sub-cohorts. The fact that this effect modification by time since exposure only becomes apparent when mixing a sub-cohort of comparatively young miners with high quality exposure assessment and a sub-cohort of comparatively older miners with low-quality exposure assessment suggests that this effect modification could be an artefact caused by exposure or dose uncertainty.

When accounting for unshared measurement error in the full cohort, we found that the correction for exposure uncertainty was only of marginal importance in the association between radon exposure and lung cancer mortality, regardless of the chosen disease model. Owing to the flexibility of the Bayesian hierarchical approach, we were able to conduct inference under a measurement model specifying a component of Berkson error shared within miners. In particular, we assumed that this shared Berkson component affected all exposure values received by a miner in a given exposure period characterised by a method of group-level exposure assessment in the same way to reflect the influence of individual job conditions and worker practices. When assuming this measurement model, we observed a marked increase in the risk coefficient associated with radon exposure, which was estimated to be 0.99 per 100 WLM. In accordance with the findings of other studies that accounted for measurement error [439, 440], we observed a widening of the 95% credible interval when accounting for shared Berkson error in the full cohort, which was estimated to be [0.48; 1.73]. Concerning the influence of effect modifying variables, the estimated risk coefficients in the disease model describing the effect modification by period of exposure were not substantially modified when accounting for unshared or shared Berkson error. In particular, irrespective of the chosen measurement model, the risk associated with exposures received after 1955 was estimated to be five times greater than the risk estimated for exposures received until 1955. Similarly, we observed an attenuation of the exposure-response relationship for high exposure values when analysing the data of the full cohort, irrespective of the chosen measurement model. We did not observe a notable increase in the risk coefficient when accounting for shared or unshared measurement error in the post-55 cohort.

If we maintain our hypothesis that the observed attenuation of the exposure-response and the observed effect modification by period of exposure are likely to be due to the differences in the methods of exposure assessment, how can we explain the fact that we still observe these phenomena when we account for measurement error?

The results of Simulation Study 2 may shed some light on this question. Indeed, the finding that we continue to observe an attenuation of the exposure-response relationship when accounting for Berkson error shared within miners could be explained by the fact that we did not account for the component that is likely to cause the most severe distortion of the exposure-response relationship in this cohort. Indeed, it is likely that we are faced with a component of classical measurement error, which is shared both within and between miners for the earliest period of exposure. In the retrospective reconstruction of the exposure values received until 1955 in the French cohort of uranium miners, a group of experts estimated only one exposure level for a given mining location. Moreover, the exposure level for a given mining location was often estimated to be the same for all years of exposure until 1955. This error-prone exposure estimate was then used to reconstruct all individual exposure values received by all miners who were employed at the given mining location until 1955. Based on this retrospective reconstruction, the precision of the exposure level estimated for a mining location can lead to a large component of classical measurement error shared both within and between miners. As we did not account for this error component in the analysis of the lung cancer mortality data of the French cohort of uranium miners, it is not surprising that we still observe an attenuation of the exposure-response relationship in the full cohort, even after accounting for shared Berkson error. As there was no retrospective exposure reconstruction in the sub-cohort of miners employed after 1955, it is not surprising either that we do not observe an attenuation of the exposure-response curve for this sub-cohort. In conclusion, it is advisable to account for this component of shared classical measurement error in future analyses of the association between radon exposure and lung cancer mortality in the French cohort of uranium miners in order to obtain reliable risk estimates in this cohort.

Finally, we calculated absorbed lung doses via the dose model specified in section 5.1.4 in the post-55 cohort. When relating lung cancer mortality in the post-55 cohort to absorbed lung dose, we found a risk coefficient of 5.92 per Gray with a credible interval of [2.48; 11.60]. This result is consistent with the findings obtained by Rage et al. (2012) [88], who found an excess relative risk of 4.59 per Gray with a 95% confidence interval of [1.31; 11.16] for radon progeny, when calculating absorbed lung doses in this sub-cohort via the Human Respiratory Tract Model (HRTM) and when modelling lung cancer mortality via Poisson regression. When accounting for exposure and dose uncertainty in the post-55 cohort by combining the disease, the measurement and the dose model in a hierarchical structure, we did not observe a substantial increase in the risk coefficient associated with absorbed lung dose.

Moreover, we accounted for dose uncertainty in the estimation of absorbed lung doses by specifying informative prior distributions on the input parameters involved in the dose model. These prior distributions were chosen in accordance with measurements performed in mines and recommendations in the literature (see section 5.1.5). It is likely that the fact that we did not observe a substantial difference in risk estimation when accounting for dose uncertainty can mainly be explained by a misspecification of priors. Indeed, even though we know, or at least we suspect, that there were systematic changes in the unknown input parameters intervening in dose calculation, the prior distributions on these input parameters do not translate this knowledge for the moment. For instance for average breathing rate, we assumed a log-normal distribution that was the same for all miners and all periods of exposure. However, it is likely that there

are big differences between different working conditions and for different exposure periods for this input parameter. To integrate this information we have established an elicitation procedure to specify an informative prior on the average breathing rate of a French uranium miner based on expert knowledge. The resulting prior distributions indicate substantial differences for the different working conditions. Contrary to the log-normal prior that we assumed for the moment, the prior distributions based on expert knowledge are more in accordance with the intervals that are recommended by the International Commission on Radiological Protection (ICRP) or the alpha risk project. In future analyses, it would be interesting to study the impact of these elicited prior distributions on average breathing rate on the estimation of the association between radon exposure and lung cancer mortality in the French cohort of uranium miners. While the three methods for the combination of expert opinion that we compared in this work, namely averaging, linear pooling and a Supra-Bayesian approach, resulted in very similar prior distributions on average breathing rate, the latter approach appears to be more flexible than the two alternative approaches. In particular, the Supra-Bayesian approach for the combination of expert opinion could be introduced in the hierarchical model we assumed when accounting for exposure and dose uncertainty in the cohort, thereby allowing to model all available information in a unique and coherent framework. Moreover, this approach of prior elicitation could be adopted to combine the information available in the literature with the information elicited through expert opinion. Finally, to improve the elicitation task we conducted in this work, we could combine the chosen indirect elicitation exercise with a dynamic graphical feedback. The application we developed allows to give a graphical feedback where experts can indicate their confidence concerning two alternative probability distributions. In our experience with this application, experts often commented their evaluations by indicating that they would have preferred more probability mass at a certain point in the probability distribution. If they were presented with an interactive graphical display of the probability distributions, similar to the histogram method described by Soares et al. (2011) [350] for instance, they would have the possibility to adjust their elicited opinion in that case, ultimately leading to a more reliable representation of their opinion. While the direct use of an interactive graphical elicitation method would be another alternative, it may be quite challenging for an expert to directly produce the entire probability distribution to translate their uncertainty on the parameter of interest. The combination of the indirect elicitation task we used in the elicitation of expert opinion with the possibility to modify the resulting probability distributions in an interactive way seems therefore a promising way to further improve the method for the elicitation of expert opinion that we employed in this work.

7.2 Perspectives

7.2.1 Implications for radiation protection

As the last uranium mine in France closed in 2001 and as uranium mining has disappeared in many other countries in which this industry was still active some 30 years ago, one may expect that the follow-up of uranium miners is of little importance for the definition of radiation protection guidelines for workers today. Quite contrary to this expectation, there has been a new

boom in uranium mining [441]. Due to an ever increasing demand for energy, and recently in particular for energy which is low in greenhouse gas emission, there has been an increasing interest in uranium exploration, particularly in developing countries [442, 443, 444]. Moreover, the precise estimates on the health effects of radon exposure are not only important for the purpose of establishing protection guidelines in underground mines, but also to protect the health of the general population. Residential radon exposure is believed to be responsible for 2% of deaths from cancer in Europe [23]. The installation of better thermal insulation in buildings to reduce the energy consumption in heating may further increase residential radon exposure, as this thermal insulation typically decreases the air exchange rates in these buildings. In order to be able to assess the cost-effectiveness of possible measures for the reduction of radon concentrations in buildings, it is indispensable to understand the health risks associated with radon exposure. In contrast to studies on uranium miners, studies on residential radon mostly use a case-control design and show conflicting results [24, 151, 23]. In light of the many sources of uncertainties which arise in the analysis of the association between radon exposure and lung cancer mortality in cohorts of uranium miners, we can ask ourselves whether these analyses are a suitable source of information to establish guidelines for radiation protection. Uranium miners are not only almost exclusively adult males [445, 90], but they are first exposed as adults. Furthermore, they may be exposed to additional potentially lung carcinogenic substances, including silica quartz [84], arsenic [446], diesel exhaust [447] and asbestos [448, 449] and present breathing characteristics which are different from those of subjects which are exposed in their homes [450].
Despite of these drawbacks, there are also various advantages of extrapolating risk estimates based on uranium miners to complement the results of case-control studies on residential radon exposure. Firstly, studies on uranium miners generally consist in large prospective cohort studies with a low percentage of subjects lost to follow-up (in the French cohort of uranium miners this percentage is less than 1%). Due to their design, analyses on cohorts of uranium miners probably neither suffer from selection bias nor recall bias for tobacco consumption, two problems which may be encountered in case-control studies on residential radon. Moreover, in studies on residential radon exposure, smoking and radon exposure tend to be negatively correlated, since in urban areas there is both a lower concentration in residential radon and a higher prevalence in smoking [23], whereas previous analyses on the impact of smoking in occupational cohort studies on uranium miners suggest that smoking is not a source of confounding in these studies [451, 452]. As the measurement error occurring in variables to assess smoking patterns in case-control studies may be quite complex and in particular differential, it is difficult to account for this type of measurement error and uncorrected risk estimates can be expected to be at best partially adjusted for smoking in these studies. Thereby, the negative correlation between radon exposure and tobacco consumption can lead to a systematic underestimation of risk coefficients associated with radon exposure in studies on residential radon. Moreover, in case-control studies on residential radon, exposure uncertainty will likely be greater than in occupational cohort studies, because the exposure assessment is mostly conducted retrospectively and few measurements are taken to reconstruct the exposure history of several decades. In occupational cohort studies, on the other hand, there is often a prospective and direct exposure assessment, at least for the more recent exposure periods. At the same time, the variability of radon exposure tends to be smaller in studies on residential radon than in occupational studies on cohorts of under-

ground miners. Note that the exposure variance in occupational cohort studies might be further amplified through the cumulation of annual exposure values. Indeed, under the assumption of a positive correlation between the annual exposure values of a miner, the variance of cumulated exposure is greater than the sum of the variances of annual exposure. Therefore this cumulation will probably make the results on the health effects of radon exposure that are based on cohorts of uranium miners even less vulnerable to the impacts of exposure uncertainty. In light of these facts, it seems important to complement the information obtained by case-control studies on residential radon by extrapolating results on the association between radon exposure and lung cancer mortality obtained on cohorts of occupationally exposed underground miners. In comparison to most other cohorts of underground miners, the French cohort of uranium miners is characterised by low levels and low rates of radon exposure for a long period of time and prospective exposure assessment of good quality, making it especially valuable for the extrapolation to residential exposure. To account for breathing characteristics, which are likely to differ between occupational and residential exposure, it is indispensable to calculate absorbed lung doses in cohorts of uranium miners and to account for all sources of uncertainty in dose calculation [453]. In accordance, the ICRP recently announced its intention to use a dosimetric approach in the study of the health effects of radon exposure that involves the calculation of radiation doses to the respiratory tract, thereby treating radon like all other radiologically relevant radioisotopes [114, 110, 454]. In this work, we showed that the Bayesian hierarchical approach can be used to account for exposure and dose uncertainty in the association between radon exposure and lung cancer mortality.

However, the findings of the simulation studies we conducted underline the importance of the correction for shared and unshared exposure uncertainty in radiation epidemiology, as they show that complex error structures in occupational cohort studies can lead to an attenuation of the exposure-response relationship for high exposure values. This phenomenon, which is frequently observed in occupational cohort studies, may pose serious challenges in risk modelling. Indeed, if this attenuation reflects the association between true exposure and the disease outcome and a linear model is chosen, it may cause a severe underestimation of risk for workers with low exposures. On the other hand, if the association between true exposure and the outcome is linear and the observed distortion of the exposure-response relationship is caused by measurement error, fitting a non-linear or a piecewise-linear model can lead to an overestimation of the risk coefficient for workers with low exposures. To inform radiation protection guidelines for workers, researchers are particularly interested in the low exposure range occurring in occupational cohort studies, because exposure levels of workers today tend to be much lower than in the past. Moreover, these exposure values are comparable to exposures received by the general population. Ignoring the cause of an observed distortion of the exposure-response curve may therefore seriously limit the extrapolability of risk estimates obtained in occupational studies in radiation epidemiology to the general population. Finally, our findings suggest that the effect modification of the association between radon exposure and lung cancer mortality by time since exposure, a well-studied phenomenon which is observed in many studies on cohorts of uranium miners [385, 384, 72, 90], may be an artefact caused by the complex patterns of exposure uncertainty in these studies. Indeed, we neither observed this effect modification for miners employed until 1955 nor for miners employed after 1955, but merely when mixing these two sub-cohorts which

present a very different quality of exposure assessments. Similarly, the results obtained by Hunter et al. (2013) show a marked effect modification by time since exposure when analysing pooled data of the three European cohorts (the Czech, the German and the French cohort of uranium miners). When restricting data to miners who were exposed to less than 50 WLM, however, this effect modification was no longer observable. Moreover, the effect modification by time since exposure is not observed in studies on residential radon. In future research, it would be interesting to study the influence of time since exposure in the Post-60 cohort of the Wismut cohort and the Post-68 cohort of the Czech cohort and in the sub-cohort of miners who were employed before these years. In that way, it would be possible to assess whether there is a genuine effect modification by time since exposure in some cohorts of uranium miners or whether this effect modification is merely an artefact that can arise when mixing a relatively young sub-cohort with good exposure assessment and a relatively older sub-cohort with poor exposure assessment, as it seems to be the case in the French cohort of uranium miners.
Finally, based on the findings on the real and on the simulated data sets concerning the impact of measurement error in cohorts of uranium miners, we would recommend to either base risk estimates on analyses which are corrected for all sources of shared and unshared exposure and dose uncertainty or to exclusively base risk estimation on those sub-cohorts of uranium miners for which exposure was assessed in a prospective fashion.

7.2.2 The use of the Bayesian hierarchical approach in radiation epidemiology

In radiation epidemiology, the use of a hierarchical approach offers the possibility to account for complex structures of exposure and dose uncertainty and to describe the different sources of uncertainty through the linking of conditional independence models. It may be challenging to conduct statistical inference for these models, because they can lead to a complex and high-dimensional likelihood and in practice it may be intractable to integrate and to maximise this high-dimensional likelihood. Moreover, in complex hierarchical models, we can often be faced with the situation where the unknown parameters in some sub-models are only poorly informed by the data and thereby require other sources of information. The Bayesian approach to conduct statistical inference offers an elegant solution to include external information through the specification of informative prior distributions and to conduct inference via MCMC methods.
As the field of radiation epidemiology is particularly interested in the estimation of the health effects of chronic low-dose exposure to ionising radiation, it is faced with many uncertainties. In particular, a great challenge in this field is the extrapolation of the observed health effects of exposures received at high to moderate doses to low-dose exposures. In the situation where it is unethical to add new study participants by deliberately exposing them to ionising radiation, we have to find other solutions to increase the statistical power and to reduce the uncertainty in studies in radiation epidemiology. The integration of external information through the specification of informative prior distributions and the modelling of all sources of uncertainty in a Bayesian hierarchical approach have the potential to increase statistical power and to reduce our uncertainty in risk estimation. Moreover, when accounting for all sources of exposure and dose uncertainty, we can hope to avoid distortions of the dose-response relationship which would make any extrapolation from populations that received high to moderate doses to inform us about the health effects of low-dose and low-dose rate exposures futile. Preston et al. (2013) [19] identified

the reliable combination of epidemiological data with experimental laboratory animal and cellular data as one important challenge in the reduction of uncertainties in estimating health risks associated with exposure to ionising radiation. Based on its coherence and its flexibility when it comes to the modelling of complex probability models, the Bayesian hierarchical approach is arguably the most promising direction to achieve the combined modelling of epidemiological and experimental data.

More generally, epidemiology is often faced with the challenge of the estimation of weak associations between an exposure of interest and a disease outcome in a context where these associations may be distorted by the possibility of complex interactions with other factors and the effects of measurement error. In order to address this challenge, it can be argued that we need to specify a hierarchical model that can describe measurement error, both in the exposure of interest and the confounding variables to avoid residual confounding. To be able to estimate the parameters in all sub-models of this hierarchical model and to be able to assure a certain statistical power when estimating weak associations, we need to be able to integrate prior knowledge based on both previous studies and on expert knowledge. Moreover, in this situation, it may simply be impossible to establish the truth of a given hypothesis based on a single well-designed epidemiological study. As pointed out by Thiebaut et al. (2007) in the context of nutritional epidemiology, "Owing to the current limitations of available procedures and reference instruments, we cannot assume that corrected estimates of diet-disease associations in any single study are definitive. To firmly establish a hypothesis, we need carefully conducted studies in diverse populations with different dietary patterns and ranges of intake, incidence rates, and sociocultural histories." [127]. To address questions that arise in epidemiology, we need an approach that can guide public health decisions. Popper's hypothetico-deductive approach is often resumed in the idea that "to assert from innumerable observations that all swans are white does not protect the affirmation against the later discovery of a non-white swan" [455]. While this idea is indubitably true, it is questionable to what extent this knowledge can help us in a context where we have to make decisions and take actions in the treatment of swans. In epidemiology, the idea is to guide decisions on public health measures and the consequences of these decisions may have very different effects in practice. In particular, to judge that the exposure to a certain substance is innocuous when it is harmful may have more serious consequences than to claim that this substance is harmful when it is innocuous. Contrary to the frequentist approach, the Bayesian approach to statistical inference has a solid decision-theoretic foundation that allows to account for different loss functions in a coherent framework [359, 292]. Finally, the Bayesian approach can provide an answer to the question that researchers are most interested in, namely "what is the probability that there is an association between an exposure and a disease outcome of interest?".

The integration of external information through the specification of informative prior distributions introduces a certain subjectivity in the analyses of epidemiological studies. Given this subjectivity, which contradicts the ideal of the strict objectivity of scientific research, is it worthwhile to introduce external information through the specification of prior distributions?

There is no more arbitrariness in the specification of a prior distribution than there is arbitrariness in the choice of a probability model. As the subjective choices involved in the specification of a probability model will in general have a greater effect on statistical inference than will

the choice of prior distributions on the unknown parameters [456, 245], it is surprising that researchers show more discomfort with the former than with the latter.

Finally, conducting Bayesian inference to estimate the parameters in a hierarchical model can be time-consuming both in the development of a problem-tailored MCMC implementation and in the estimation process. Given this effort, is the use of a Bayesian hierarchical approach to account for exposure and dose uncertainty suitable for the analysis of the health effects in epidemiological studies?

An alternative can be to spent large amounts of money and time in the design and in the analysis of animal studies to find the biological basis of effect modifying variables, which may be only an artefact caused by measurement error. Moreover, compared to the years it often takes to set up an epidemiological study and to organise its follow-up, it can be argued that the time spent in the development and especially in the computation of risk estimates that are corrected for exposure and dose uncertainty is negligible.

7.2.3 Overcoming challenges in the implementation of a Bayesian hierarchical approach

Based on the points we discussed so far, the Bayesian hierarchical approach seems to be a promising option to address many challenges in epidemiology. However, the implementation of a Bayesian hierarchical approach involves two main obstacles: The first is the construction of a hierarchical model which involves the specification of all sub-models to describe different error structures and other sources of uncertainty. The second is how to obtain and to interpret estimates of the parameters of interest. Based on these obstacles, we would not fully agree with Adam Smith, who is reported to have said that with the combination of the Bayesian paradigm and MCMC all problems of statistics had been solved. According to him, "there was nothing else to do with statistical problems but to plug them into a computer and turn the Bayesian crank" [224].

The first obstacle can be easily overcome by interdisciplinary collaborations between statisticians and researchers in radiation epidemiology and in other domains. The experiences of the specialists in the different domains can provide invaluable information, both when it comes to the definition of the different sub-models involved in the hierarchical model and in the specification of prior distributions. In view of data structures and problems addressed in epidemiology that are becoming ever more complex, interdisciplinary collaborations are both becoming indispensable and providing considerable contributions to the quality of the studies that arise from these collaborations.

The main obstacle in the idea to plug statistical problems into a computer and to "turn the Bayesian crank" is that there are so far no statistical software packages that would provide the possibility for the fitting of the complex hierarchical models that are necessary to account for exposure uncertainty in epidemiological studies in a time-efficient manner. In many cases, we still need to develop a problem-tailored MCMC implementation to conduct Bayesian inference for the specified hierarchical model to address challenges like the cumulation of individual exposure values to obtain cumulated exposure or to accelerate the updating of the high-dimensional unknown quantities. Although they convincingly demonstrated the advantages of a Bayesian

approach to conduct statistical inference, the heroes of the glorious Bayesian revival, including Jeffreys, Savage and Good, were not able to convince the scientific community to actually use this approach. Similarly, we should not have too many illusions about the extent to which the hierarchical approach to account for exposure and dose uncertainty will be used in radiation epidemiology or in other domains unless they exist software packages that provide an efficient MCMC implementation to conduct Bayesian inference in this situation.
The software application for the elicitation of prior information based on expert knowledge that we developed in this work can be seen as a contribution to overcome both obstacles. Firstly, it provides a generic software implementation that can be applied in other studies and, secondly, the elicitation of informative prior distributions brings statisticians closer to their subject-matter colleagues [347]. Several authors have argued that the elicitation of prior distributions is an important part of understanding the parameters in a probability model [347, 343]. This understanding is a necessary step to be able to interpret the posterior distribution [343] and more generally to provide answers that applied researchers are interested in.

7.3 Conclusion

To conclude, the hierarchical approach, combined with the Bayesian framework to conduct statistical inference, appears to be an extremely powerful and promising alliance in the study of the health effects of radiation exposure in epidemiological studies. In radiation epidemiology, it is of great importance to account for uncertainties and to integrate external information in the analysis of epidemiological studies to obtain more reliable and more precise estimates when modelling the health effects of radiation exposure. The Bayesian hierarchical approach provides a promising solution to both challenges by accounting for all available information and all sources of uncertainty in a unique and coherent framework. Moreover, it can provide answers to questions that are of interest to researchers, radiation protection committees and to the general public. Nonetheless, the use of the Bayesian hierarchical approach to account for exposure and dose uncertainty in epidemiological studies requires the development of generic software packages that allow to use this approach in a broad range of problems and it makes interdisciplinary collaborations indispensable. Finally, with great power comes great responsibility, and the Bayesian hierarchical approach can neither dispense with the parsimonious modelling of the observed phenomena nor with model checking and sensitivity analyses concerning the specified prior distributions.

Bibliography

1 Abbas AE. A comparison of two probability encoding methods: fixed probability vs. fixed variable values. Decision Analysis. 2008;5(4):190–202.

2 Birchall A, James AC. Uncertainty analysis of the effective dose per unit exposure from radon progeny and implications for ICRP risk-weighting factors. Radiation Protection Dosimetry. 1994;53(1-4):133–140.

3 Thomas D, Stram D, Dwyer H. Exposure measurement error: influence on exposure-disease relationships and methods of correction. Annual Review of Public Health. 1993;14(1):69–93.

4 Kim HM, Yasui Y, Burstyn I. Attenuation in risk estimates in logistic and Cox proportional-hazards models due to group-based exposure assessment strategy. Annals of Occupational Hygiene. 2006;50(6):623–635.

5 Physick LW, Cope ME, Lee S, Hurley PJ. An approach for estimating exposure to ambient concentrations. Journal of Exposure Science and Environmental Epidemiology. 2007;17:76–83.

6 Carroll RJ. Measurement error in epidemiologic studies. In: Encyclopedia of biostatistics. vol. 5. Wiley Online Library; 2005. .

7 Blair A, Thomas K, Coble J, Sandler DP, Hines CJ, Lynch CF, et al. Impact of pesticide exposure misclassification on estimates of relative risks in the Agricultural Health Study. Occupational and Environmental Medicine. 2011;68(7):537–41.

8 Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. Boca Raton: Chapman Hall; 2006.

9 Richardson S, Leblond L. Some comments on misspecification of priors in Bayesian modelling of measurement error problems. Statistics in Medicine. 1997;16:203–213.

10 de Dieu Tapsoba J, Lee SM, Wang CY. Expected estimating equation using calibration data for generalized linear models with a mixture of Berkson and classical errors in covariates. Statistics in Medicine. 2013;33(4):675–692.

11 Perrier F, Giorgis-Allemand L, Slama R, Philippat C. Within-subject pooling of biological samples to reduce exposure misclassification in biomarker-based studies. Epidemiology. 2016;27(3):378–388.

12 Jurek AM, Maldonado G, Greenland S, Church TR. Exposure-measurement error is frequently ignored when interpreting epidemiologic study results. European Journal of Epidemiology. 2006;21:871–876.

13 Spiegelman D. Approaches to uncertainty in exposure assessment in environmental epidemiology. Annual Review of Public Health. 2010;31:149–63.

14 Steenland K, Karnes C, Darrow L, Barry V. Attenuation of exposure-response rate ratios at higher exposures: A simulation study focusing on frailty and measurement error. Epidemiology. 2015;26(3):395–401.

15 Greenland S, Fischer HJ, Kheifets L. Methods to Explore Uncertainty and Bias Introduced by Job Exposure Matrices. Risk Analysis. 2016;36(1):74–82.

16 Kromhout H. Design of measurement strategies for workplace exposures. Occupational and Environmental Medicine. 2002;59:349 – 354.

17 Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scandinavian Journal of Work, Environment & Health. 2003;29:317–324.

18 Blair A, Stewart P, Lubin JH, Forastiere F. Methodological issues regarding confounding and exposure misclassification in epidemiological studies of occupational exposures. American Journal of Industrial Medicine. 2007;50(3):199–207.

19 Preston RJ, Boice JD Jr, Brill AB, Chakraborty R, Conolly R, Hoffman FO, et al. Uncertainties in estimating health risks associated with exposure to ionising radiation. Journal of Radiological Protection. 2013;33(3):573–88.

20 Birchall A, Marsh JW. Radon dosimetry and its implication for risk. International Congress Series. 2005;p. 81–84.

21 IARC WG. Man-made mineral fibres and radon / IARC Working Group on the Evaluation of Carcinogenic Risks to Humans which met in Lyon, 16-23 June 1987. IARC Lyon; 1988.

22 Samet JM, Eradze GR. Radon and lung cancer risk: taking stock at the millenium. Environmental Health Perspectives. 2000;108(Suppl 4):635–641.

23 Darby S, Hill D, Auvinen A, Barros-Dios J, Baysson H, Bochicchio F, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ. 2005;330(7485):223.

24 Lubin JH, Boice Jr JD, Samet JM. Errors in exposure assessment, statistical power and the interpretation of residential radon studies. Radiation Research. 1995;144(3):329–341.

25 Stram D, Langholz B, Huberman M, Thomas D. Correcting for exposure measurement error in a reanalysis of lung cancer mortality for the Colorad Plateau uranium miners cohort. Health Physics. 1999;77(3).

26 Reeves GK, Cox DR, Darby SC, Whitley E. Some aspects of measurement error in explanatory variables for continuous and binary regression models. Statistics in Medicine. 1998;17:2157 –2177.

27 Darby S, Whitley E, Silcocks P, Thakrar B, Green M, Lomas P, et al. Risk of lung cancer associated with residential radon exposure in south-west England: a case-control study. British Journal of Cancer. 1998;78(3).

28 Heid I, Küchenhoff H, Wellmann J, Gerken M, Kreienbrock L, Wichmann HE. On the potential of measurement error to induce differential bias on odds ratio estimates: an example from radon epidemiology. Statistics in Medicine. 2002;21:3261–3278.

29 Heid IM, Küchenhoff H, Miles J, Kreienbrock L, Wichmann HE. Two dimensions of measurement error: Classical and Berkson error in residential radon exposure assessment. Journal of Exposure Analysis and Environmental Epidemiology. 2004;14:365–377.

30 Heidenreich WF, Luebeck EG, Moolgavkar SH. Effects of exposure uncertainties in the TSCE model and application to the Colorado miners data. Radiation Research. 2004;161(1):72–81.

31 Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. Statistics in Medicine. 2005;24(11):1713 – 1723.

32 Küchenhoff H, Bender R, Langner I. Effect of Berkson measurement error on parameter estimates in Cox regression models. Lifetime Data Analysis. 2007;13(2):261–272.

33 Allodji RS, Leuraud K, Bernhard S, Henry S, Bénichou J, Laurier D. Assessment of uncertainty associated with measuring exposure to radon and decay products in the French uranium miners cohort. Journal of Radiological Protection. 2012;32(1):85–100.

34 Allodji RS, Leuraud K, Thiébaut AC, Henry S, Laurier D, Bénichou J. Impact of measurement error in radon exposure on the estimated excess relative risk of lung cancer death in a simulated study based on the French Uranium Miners' Cohort. Radiation and Environmental Biophysics. 2012;51(2):151–163.

35 Allodji RS, Thiébaut A, Leuraud K, Rage E, Henry S, Laurier D, et al. The performance of functional methods for correcting non-Gaussian measurement error within Poisson regression: corrected excess risk of lung cancer mortality in relation to radon exposure among French uranium miners. Statistics in Medicine. 2012;31(30):4428–4443.

36 Heidenreich W, Tomasek L, Grosche B, Leuraud K, Laurier D. Lung cancer mortality in the European uranium miners cohorts analyzed with a biologically based model taking into account radon measurement error. Radiation and Environmental Biophysics. 2012;51(3):263–275.

37 Marsh JW, Birchall A. Sensitivity analysis of the weighted equivalent lung dose per unit exposure from radon progeny. Radiation Protection Dosimetry. 2000;87(3):167–178.

38 Marsh JW, Birchall A, Butterweck G, Dorrian MD, Huet C, Ortega X, et al. Uncertainty analysis of the weighted equivalent lung dose per unit exposure to radon progeny in the home. Radiation Protection Dosimetry. 2002;102(3):229–48.

39 Marsh JW, Birchall A. Uncertainty analysis of the absorbed dose to regions of the lung per unit exposure to radon progeny in a mine. Health Protection Agency; 2009.

40 Advisory Group on Ionising Radiation AGIR. Radon and Public Health. Health Protection Authority; 2009.

41 George AC, Paschoa AS, Steinhäusler F. World hisotry of radon research and measurement from the early 1900's to today. In: AIP Conference Proceedings. vol. 1034. AIP; 2008. p. 20–33.

42 Robinson RF. Mining and selling radium and uranium. Springer International Publishing; 2015.

43 Saccomanno G, Archer VE, Saunders RP, James LA, Beckler PA. Lung cancer of uranium miners on the colorado plateau. Health Physics. 1964;10:1195–1201.

44 Porstendörfer J. Properties and behaviour of radon and thoron and their decay products in the air. Journal of Aerosol Science. 1994;25(2):219–263.

45 Bale WF. Memorandum to the Files, March 14, 1951: Hazards associated with radon and thoron. Health Physics. 1980;38(6):1062–1066.

46 Saccomanno G, Huth GC, Auerbach O, Kuschner M. Relationship of radioactive radon daughters and cigarette smoking in the genesis of lung cancer in uranium miners. Cancer. 1988;62(7):1402–1408.

47 Tirmarche M, Harrison J, Laurier D, Blanchardon E, Paquet F, Marsh J. Risk of lung cancer from radon exposure: contribution of recently published studies of uranium miners. Annals of the ICRP. 2012;368–377(3):368–377.

48 Rhodes R. The making of the atomic bomb. Simon & Schuster; 1987.

49 Ferronsky VI. Nuclear Geophysics - Applications to Hydrology, Hydrogeology, Engineering Geology, Agriculture and Environmental Science. Springer International Publishing; 2015.

50 Domenech H. Radiation Safety - Management and Programs. Springer International Publishing; 2017.

51 Pathways to Modern Chemical Physics. Springer Berlin Heidelberg; 2012.

52 Basdevant JL, Rich J, Spiro M. Fundamentals in nuclear physics. Springer; 2005.

53 Gooch JW. Encyclopedic dictionary of polymers. Springer Science + Business Media; 2007.

54 Jevremovic T. Nuclear principles in engineering. Springer US; 2005.

55 Fényes T. Basic properties of the atomic nucleus. In: Vértes A, Nagy S, Klencsár Z, Lovas RG, Rösch F, editors. Handbook of nuclear chemistry. Springer Science + Business Media; 2011. p. 39–141.

56 Nature's building blocks. Oxford University Press; 2001.

57 Griffin HC. Natural Radioactive Decay Chains. In: Handbook of nuclear chemistry. Attila Vértes and Sándor Nagy and Zoltán Klencsár and Rezso G Lovas and Frank Rösch; 2011. .

58 Lecomte J. Radon and the system of radiological protection. Annals of the ICRP. 2011;41(3):389–396.

59 Marsh JW, Harrison JD, Laurier D, Blanchardon E, Paquet F, Tirmarche M. Dose conversion factors for radon: recent developments. Health Physics. 2010;99(4):511–516.

60 Marsh JW, Blanchardon E, Gregoratto D, Hofmann W, Karcher K, Nosske D, et al. Dosimetric calculations for uranium miners for epidemiological studies. Radiation Protection Dosimetry. 2012;149(4):371–383.

61 Jelle BP. Development of a model for radon concentration in indoor air. Science of the Total Environment. 2012;416:343–50.

62 Lubin JH, Boice JD, Edling C, Hornung RW, Howe GR, Kunz E, et al. Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. Journal of the National Cancer Institute. 1995;87(11):817–827.

63 Beir, VI. Health effect of exposure to radon. Comittee on health risks of exposure to radon, Board on radiation effects research, Commission on life sciences, National Research council. 1999;.

64 Laurier D, Tirmarche M, Mitton N, Valenty M, Richard P, Poveda S, et al. An update of cancer mortality among the French cohort of uranium miners: extended follow-up and new source of data for causes of death. European Journal of Epidemiology. 2004;19(2):139–146.

65 Hatton F, Bouvier-Colle M, Michel E, Maujol L. Les causes de décès en France. In: Bouvier-Colle M, Vallin J, Hatton F, editors. Mortalité et causes de décès en France. Paris: Les Editions INSERM; 1990. p. 111–145.

66 Lubin JH, Boice Jr JD, Edling C, Hornung RW, Howe G, Kunz E, et al. Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. Health Physics. 1995;69(4):494–500.

67 Waggit P. Uranium mining legacies remediation and renaissance development: an international overview. In: Merkel BJ, Hasche-Berger A, editors. Uranium, mining and hydrogeology. Springer Science + Business Media; 2008. p. 11–18.

68 Lane RS, Frost SE, Howe GR, Zablotska LB. Mortality (1950-1999) and cancer incidence (1969-1999) in the cohort of Eldorado uranium workers. Radiation Research. 2010;174(6a):773–785.

69 Marušiaková M, Gregor Z, Tomášek L. A review of exposures to radon, long-lived radionuclides and external gamma at the Czech uranium mine. Radiation Protection Dosimetry. 2011;145(2-3):248–51.

70 Vacquier B, Rogel A, Leuraud K, Caer S, Acker A, Laurier D. Radon-associated lung cancer risk among French uranium miners: modifying factors of the exposure–risk relationship. Radiation and Environmental Biophysics. 2008;48(1):1–9.

71 Kreuzer M, Schnelzer M, Tschense A, Walsh L, Grosche B. Cohort profile: the German uranium miners cohort study (WISMUT cohort), 1946-2003. International Journal of Epidemiology. 2010;39(4):980–7.

72 Tomasek L, Agnes R, Tirmarche M, Mitton N, Laurier D. Lung cancer in French and Czech uranium miners: radon-associated risk at low exposure rates and modifying effects of time since exposure and age at exposure. Radiation Research. 2008;169(2):125–137.

73 Navarajan G, Berriault C, Do M, Villeneuve P, Demers PA. Cancer incidence and mortality from exposure to radon progeny among Ontario uranium miners. Occupational and Environmental Medicine. 2016;73:838–845.

74 Tirmarche M, Raphalen A, Allin F, Chameaud J, Bredon P. Mortality of a cohort of French uranium miners exposed to relatively low radon concentrations. British Journal of Cancer. 1993;67:1090–1097.

75 Tirmarche M. The present state of an epidemiological study of uranium miners in France. In: Occupational radiation safety in mining; 1984. .

76 Fischer HJ, Vergara XP, Yost M, Silva M, Lombardi DA, Kheifets L. Developing a job-exposure matrix with exposure uncertainty from expert elicitation and data modeling. Journal of Exposure Science and Environmental Epidemiology. 2017;27(1):7–15.

77 Rogel A, Laurier D, Tirmarche M, Quesne B. Lung cancer risk in the French cohort of uranium miners. Journal of Radiological Protection. 2002;22(3A):A101.

78 Zettwoog P. State-of-the-art of the *alpha* individual dosimetry in France. In: Gomez M, editor. Radiation hazards in mining: Control, measurements and medical aspects; 1981. p. 4–9.

79 George AC. The history, development and the present status of the radon measurement programme in the United States of America. Radiation Protection Dosimetry. 2015;167(1-3):8–14.

80 Tomasek L. Lung cancer mortality among Czech uranium miners – 60 years since exposure. Journal of Radiological Protection. 2012;32(3):301–314.

81 Rage E, Caër-Lorho S, Drubay D, Ancelet S, Laroche P, Laurier D. Mortality analysis in the updated French cohort of uranium miners (1946 – 2007). International Archives of Occupational and Environmental Health. 2015;88(6):717–730.

82 Loomis D, Richardson DB, Elliott L. Poisson regression analysis of ungrouped data. Occupational and Environmental Medicine. 2005;62(5):325–9.

83 Council NR. Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII-Phase 2. Washington, DC; 2005.

84 Schubauer-Berigan MK, Daniels RD, Pinkerton LE. Radon exposure and mortality among white and American Indian uranium miners: an update of the Colorado Plateau cohort. American Journal of Epidemiology. 2009;169(6):718–730.

85 Kreuzer M, Sobotzki C, Fenske N, Marsh JW, Schnelzer M. Leukaemia mortality and low-dose ionising radiation in the WISMUT uranium miner cohort (1946–2013). Occupational and Environmental Medicine. 2017;74:252–258.

86 Walsh L, Dufey F, Tschense A, Schnelzer M, Grosche B, Kreuzer M. Radon and the risk of cancer mortality - Internal Poisson models for the German uranium miners cohort. Health Physics. 2010;99(3).

87 Kreuzer M, Fenske N, Schnelzer M, Walsh L. Lung cancer risk at low radon exposure rates in German uranium miners. British Journal of Cancer. 2015;113(9):1367–1369.

88 Rage E, Vacquier B, Blanchardon E, Allodji RS, Marsh JW, Caër-Lorho S, et al. Risk of lung cancer mortality in relation to lung doses among French uranium miners: follow-up 1956-1999. Radiation Research. 2012;177:288–297.

89 Hornung RW, Deddens J, Roscoe R. Modifiers of exposure-response estimates for lung cancer among miners exposed to radon progeny. Environmental Health Perspectives. 1995;103 Suppl 2:49–53.

90 Hunter N, Muirhead CR, Tomasek L, Kreuzer M, Laurier D, Leuraud K, et al. Joint analysis of three European nested case-control studies of lung cancer among radon exposed miners: exposure restricted to below 300 WLM. Health Physics. 2013;104(3):282–292.

91 Kreuzer M, Sobotzki C, Schnelzer M, Fenske N. Factors modifying the radon-related lung cancer risk at low exposures and exposure rates among German uranium miners. Radiation Research. 2017;.

92 Tomásek L. Czech miner studies of lung cancer risk from radon. Journal of Radiological Protection. 2002;22(3A):A107–12.

93 Leuraud K, Billon S, Bergot D, Tirmarche M, Caër S, Quesne B, et al. Lung cancer risk associated to exposure to radon and smoking in a case-control study of French uranium miners. Health Physics. 2007;92(4):371–378.

94 Leuraud K, Schnelzer M, Tomasek L, Hunter N, Timarche M, Grosche B, et al. Radon, smoking and lung cancer risk: results of a joint analysis of three European case-control studies among uranium miners. Radiation Research. 2011;176(3):375–387.

95 Howe GR, Stager RH. Risk of lung cancer mortality after exposure to radon decay products in the Beaverlodge cohort based on revised exposure estimates. Radiation Research. 1996;146(1):37–42.

96 Tomásek L, Darby SC, Fearn T, Swerdlow AJ, Placek V, Kunz E. Patterns of lung cancer mortality among uranium miners in West Bohemia with varying rates of exposure to radon and its progeny. Radiation Research. 1994;137(2):251–61.

97 Morrison HI, Villeneuve PJ, Lubin JH, Schaubel DE. Radon-progeny exposure and lung cancer risk in a cohort of newfoundland fluorspar miners. Radiation Research. 1998;150:58–65.

98 Kanyár B, Köteles GJ. Dosimetry and Biological Effects of Ionizing Radiation Dosimetry and biological effects of ionizing radiation. In: Handbook of nuclear chemistry. Attila Vértes and Sándor Nagy and Zoltán Klencsár and Rezso G Lovas and Frank Rösch; 2011. p. 2215–2257.

99 on Radiation Units TIC, Measurements. Measurement and reporting of radon exposures. ICRU; 2015.

100 ICRP103. The 2007 recommendations of the International Commission of Radiological Protection. ICRP Publication 103, Ann ICRP. 2007;.

101 el Hussein A, Ahmed AA, Mohammed A. Radiation dose to the human respiratory tract from inhalation of radon-222 and its progeny. Applied Radiation and Isotopes. 1998;49(7):783–90.

102 66 I. Human respiratory tract model for radiological protection. ICRP Publication 66, Ann ICRP. 1994;24(1-3).

103 Bailey M, Ansoborlo E, Etherington G, Gregoratto D, Guilmette R, Marsh J, et al. Proposed updating of the ICRP human respiratory tract model. In: 12th international congress of the International Radiation Protection Association (IRPA), Buenos Aires; 2008. p. 19–24.

104 Marsh JW, Harrison JD, Laurier D, Birchall A, Blanchardon E, Paquet F, et al. Doses and lung cancer risks from exposure to radon and plutonium. International Journal of Radiation Biology. 2014;90(11):1080–1087.

105 Gilbert ES. The impact of dosimetry uncertainties on dose-response analyses. Health Physics. 2009;97(5):487–92.

106 Little MP, Kwon D, Doi K, Simon SL, Preston DL, Doody MM, et al. Association of chromosome translocation rate with low dose occupational radiation exposures in U.S. radiologic technologists. Radiation Research. 2014;182.

107 Drozdovitch V, Minenko V, Golovanov I, Khrutchinsky A, Kukhta T, Kutsen S, et al. Thyroid Dose Estimates for a Cohort of Belarusian Children Exposed to (131)I from the Chernobyl Accident: Assessment of Uncertainties. Radiation Research. 2015;184(2):203–18.

108 Harley NH, Cohen BS, Robbins ES. The variability in radon decay product bronchial dose. Environment International. 1996;22:959–964.

109 Winkler-Heil R, Hofmann W, Marsh J, Birchall A. Comparison of radon lung dosimetry models for the estimation of dose uncertainties. Radiation Protection Dosimetry. 2007;127(1-4):27–30.

110 Müller WU, Giussani A, Rühm W, Lecomte JF, Harrison J, Kreuzer M, et al. Current knowledge on radon risk: implications for practical radiation protection? radon workshop, 1/2 December 2015, Bonn, BMUB (Bundesministerium für Umwelt, Naturschutz, Bau un Reaktorsicherheit; Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety). Radiation and Environmental Biophysics. 2016;55:267–280.

111 Nosske D, Blanchardon E, Bolch WE, Breustedt B, Eckerman KF, Giussani A, et al. New developments in internal dosimetry models. Radiation Protection Dosimetry. 2011;144(1-4):314–20.

112 Archer VE, Gillam JD, Wagoner JK. Respiratory disease mortality among uranium miners. Annals of the New York Academy of Sciences. 1976;271:280–93.

113 Markovic VM, Krstic D, Nikezic D. Gamma and beta doses in human organs due to radon progeny in human lung. Radiation Protection Dosimetry. 2009;135(3):197–202.

114 115 I. Lung cancer risk from radon and progeny and statement on radon. ICRP Publication 115, Annals of the ICRP. 2010;40(1).

115 Butterweck G, Porstendörfer J, Reineking A, J K. Unattached fraction and the aerosol size distribution of the radon progeny in a natural cave and mine atmospheres. Radiation Protection Dosimetry. 1992;45:167–170.

116 Marsh JW, Bessa Y, Birchall A, Blanchardon E, Hofmann W, Nosske D, et al. Dosimetric models used in the Alpha-Risk project to quantify exposure of uranium miners to radon gas and its progeny. Radiation Protection Dosimetry. 2008;130(1):101–6.

117 Broadbent A. Conceptual and methodological issues in epidemiology: An overview. Preventive Medicine. 2001;53(4–5):215–216.

118 Stram DO, Preston DL, Sokolnikov M, Napier B, Kopecky KJ, Boice J, et al. Shared dosimetry error in epidemiological dose-response analyses. PLoS One. 2015;10(3):e0119418.

119 Simon SL, Hoffman FO, Hofer E. The two-dimensional Monte Carlo: a new methodological paradigm for dose reconstruction for epidemiological research. Radiation Research. 2015;183:27–41.

120 Pan W, Zeng D, Lin X. Estimation in semiparametric transition measurement error models for longitudinal data. Biometrics. 2009;65(3):728–736.

121 Michels KB. A renaissance for measurement error. International Journal of Epidemiology. 2001;30:421–422.

122 Michels KB, Bingham SA, Luben R, Welch AA, Day NE. The effect of correlated measurement error in multivariate models of diet. American Journal of Epidemiology. 2004;160(1):59–67.

123 Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Statistics in Medicine. 1989;8:1051–1069.

124 Kipnis V, Carroll RJ, Freedman LS, Li L. Implications of a new dietary measurement error model for estimation of relative risk: Application to four calibration studies. American Journal of Epidemiology. 1999;150:642–651.

125 Kipnis V, Midthune D, Freedman LS, Bingham S, Schatzkin A, Subar A, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. American Journal of Epidemiology. 2001;153(4):394–403.

126 Day NE, Wong MY, Bingham S, Khaw KT, Luben R, Michels KB, et al. Correlated measurement error - implications for nutritional epidemiology. International Journal of Epidemiology. 2004;33:1272–1281.

127 Thiébaut ACM, Freedman LS, Carroll RJ, Kipnis V. Is it necessary to correct for measurement error in nutritional epidemiology? Annals Internal Medicine. 2007;146(1):65–7.

128 Thoresen M, Laake P. On the simple linear regression model with correlated measurement errors. Journal of Statistical Planning and Inference. 2007;137:68–78.

129 Kipnis V, Freedman LS. Impact of exposure measurement error in nutritional epidemiology. Journal of the National Cancer Institute. 2008;100(23):1658–1659.

130 Rosner B, Michels KB, Chen YH, Day NE. Measurement error correction for nutritional exposures with correlated measurement error: Use of the method of triads in a longitudinal setting. Statistics in Medicine. 2008;27(18).

131 Augustin T, Döring A, Rummel D. Regression calibration for Cox regression under heteroscedastic measurement error-Determining risk factors of cardiovascular diseases from error-prone nutritional replication data. In: Recent advances in linear models and related areas. Springer; 2008. p. 253–278.

132 Wang CY. Non-parametric maximum likelihood estimation for cox regression with subject-specific measurement error. Scandinavian Journal of Statistics. 2008;35:613–628.

133 Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, et al. Modeling data with excess zeros and measurement error: Application to evaluating relationships between episodically consumed foods and health outcomes. Biometrics. 2009;65(4):1003–1010.

134 Buonaccorsi JP, Dalen I, Laake P, Hjartaker A, Engeset D, Thoresen M. Sensitivity of regression calibration to non-perfect validation data with application to the Norwegian Women and Cancer Study. Statistics in Medicine. 2015;34.

135 Stram DO, Kopecky KJ. Power and uncertainty analysis of epidemiological studies of radiation-related disease risk in which dose estimates are based on a complex dosimetry system: Some observations. Radiation Research. 2003;160(4):408–417.

136 Bennett J, Little MP, Richardson S. Flexible dose-response models for Japanese atomic bomb survivor data: Bayesian estimation and prediction of cancer risk. Radiation and Environmental Biophysics. 2004;43(4):233–245.

137 Schafer DW, Gilbert ES. Some statistical implications of dose uncertainty in radiation dose-response analyses. Radiation Research. 2006;166(1 Pt 2):303–12.

138 Kwon D, Hoffman FO, Moroz BE, Simon SL. Bayesian dose-response analysis for epidemiological studies with complex uncertainty in dose estimation. Statistics in Medicine. 2016;35(3):399–423.

139 Preston DL, Stram DO. The growth of biostatistics and estimation of cancer risk estimates: Past, current and future challenges. Radiation Protection Dosimetry. 2017;173(1):32–35.

140 Holliday KM, Avery CL, Poole C, McGraw K, Williams R, Liao D, et al. Estimating personal exposures from ambient air pollution measures: using meta-analysis to assess measurement error. Epidemiology. 2014;25(1):35–43.

141 Hsu SOI, Ito K, Lippmann M. Effects of thoracic and fine PM and their components on heart rate and pulmonary function in COPD patients. Journal of Exposure Science and Environmental Epidemiology. 2011;21(5):464–72.

142 Suh HH, Zanobetti A. Exposure error masks the relationship between traffic-related air pollution and heart rate variability. Journal of Occupational and Environmental Medicine. 2010;52(7):685–92.

143 Reginatto M. Bayesian approach for quantifying the uncertainty of neutron doses derived from spectrometric measurements. Radiation Protection Dosimetry. 2006;121(1):64–9.

144 Bartlett JW, Keogh RH. Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration. Statistical Methods in Medical Research. 2016;.

145 Fuller WA. Measurement error models. John Wiley & Sons; 1987.

146 Biewen E, Nolte S, Rosemann M. Perturbation by multiplicative noise and the simulation extrapolation method. Advances in Statistical Analysis. 2008;92:375–389.

147 Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occupational and Environmental Medicine. 1998;55(10):651–656.

148 Gustafson P. Measurement error and misclassification in statistics and epidemiology - Impacts and Bayesian adjustments. Chapman & Hall/CRC; 2004.

149 Majumdar A. Maximum likelihood estimation of measurement error models based on the Monte Carlo EM algorithm. State University of New York at Buffalo; 2007.

150 Pierce DA, Stram DO, Vaeth M. Allowing for Random Errors in Radiation Dose Estimates for the Atomic Bomb SurvivorData. Radiation Research. 1990;123:275 – 284.

151 Lubin JH, Wang ZY, Wang LD, Boice JD Jr, Cui HX, Zhang SR, et al. Adjusting lung cancer risks for temporal and spatial variations in radon concentration in dwellings in Gansu Province, China. Radiation Research. 2005;163(5):571–9.

152 Lyles R, Kupper L. A detailed evaluation of adjustement methods for multiplicative measurement error in linear regression with applications in occupational epidemiology. Biometrics. 1997;53(3):1008–1025.

153 Zhang Z, Preston DL, Sokolnikov M, Napier BA, Degteva M, Moroz B, et al. Correction of confidence intervals in excess relative risk models using Monte Carlo dosimetry systems with shared errors. PLoS One. 2017;12(4).

154 Berkson J. Are there two regressions? Journal of the American Statistical Association. 1950;45(164–180).

155 Kim HM, Richardson D, Loomis D, Van Tongeren M, Burstyn I. Bias in the estimation of exposure effects with individual- or group-based exposure assessment. Journal of Exposure Science and Environmental Epidemiology. 2011;21(2):212–21.

156 Althubaiti A, Donev A. Non-Gaussian Berkson errors in bioassay data. Statistical Methods in Medical Research. 2016;25(1):430–445.

157 Hoffmann S, Rage E, Laurier D, Laroche P, Guihenneuc C, Ancelet S. Accounting for Berkson and classical measurement error in radon exposure using a Bayesian structural approach in the analysis of lung cancer mortality in the French cohort of uranium miners. Radiation Research. 2017;187(2):196–209.

158 Mallick B, Hoffman FO, Carroll RJ. Semiparametric regression modeling with mixtures of Berkson and classical error, with application to fallout from the Nevada test site. Biometrics. 2002;58(1):13 – 20.

159 Ahrens W, Pigeot I. Handbook of epidemiology. New York: Springer; 2005.

160 Buzas JS, Stefanski LA, Tosteson TD. Measurement error. In: Handbook of epidemiology; 2014. p. 1241–1282.

161 Guolo A. Robust techniques for measurement error correction: a review. Statistical Methods in Medical Research. 2008;17(6):555–80.

162 Buonaccorsi JP. Measurement error - Models, methods and applications. Chapman Hall/CRC; 2010.

163 Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. Sports Medicine. 1998;26(4):217–38.

164 Espino-Hernandez G, Gustafson P, Burstyn I. Bayesian adjustment for measurement error in continuous exposures in an individually matched case-control study. BMC Medical Research Methodology. 2011;11(67).

165 Li Y, Guolo A, Hoffman FO, Carroll RJ. Shared uncertainty in measurement error problems, with application to Nevada Test Site fallout data. Biometrics. 2007;63(4):1226–36.

166 Buonaccorsi JP, Lin CD. Berkson measurement error in designed repeated measures studies with random coefficients. Journal of Statistical Planning and Inference. 2002;104:53–72.

167 Spiegelman D, Valanis B. Correcting for bias in relative risk estimates due to exposure measurement error: a case study of occupational exposure to antineoplastics in pharmacists. American Journal of Public Health. 1998;88(3):406–412.

168 Hughes MD. Regression dilution in the proportional hazards model. Biometrics. 1993;49:1056–1066.

169 Carroll RJ, Stefanski LA. Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. Statistics in Medicine. 1994;13(12):1265–82.

170 Greenland S. Causation and causal inference. In: Lovric M, editor. International Encyclopedia of Statistical Science. Springer-Verlag Berlin Heidelberg; 2011. p. 216–221.

171 Thomas DC. Some Contributions of Statistics to Environmental Epidemiology. Journal of the American Statistical Association. 2000;95(449):315–319.

172 Lindqvist PG, Epstein E, Landin-Olsson M, Ingvar C, Nielsen K, Stenbeck M, et al. Avoidance of sun exposure is a risk factor for all-cause mortality: results from the Melanoma in Southern Sweden cohort. Journal of Internal Medicine. 2014;276(1):77–86.

173 Lindqvist PG, Olsson H. Answer to IM-16-0459. Journal of Internal Medicine. 2016;.

174 Thomas D. New techniques for the analysis of cohort studies. Epidemiologic Reviews. 1998;20(1):122–134.

175 Tielemans E, Kupper L, Kromhout H, Heederik D, Houba R. Individual-based and group-based occupational exposure assessment: Some equations to evaluate different strategies. Annals of Occupational Hygiene. 1998;42(2):115–119.

176 Fraser GE, Stram DO. Regression calibration in studies with correlated variables measured with error. American Journal of Epidemiology. 2001;154(9):836–844.

177 Jurek AM, Greenland S, Maldonado G, Church TR. Proper interpretation of non-differential misclassification effects: expectations vs observations. International Journal of Epidemiology. 2005;34(3):680–7.

178 Jurek AM, Greenland S, Maldonado G. How far from non-differential does exposure or disease misclassification have to be to bias measures of association away from the null? International Journal of Epidemiology. 2008;37(2):382–5.

179 Burstyn I, Yang Y, Schnatter AR. Effects of non-differential exposure misclassification on false conclusions in hypothesis-generating studies. International Journal of Environmental Research and Public Health. 2014;11(10):10951–66.

180 Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 1972;34(2):187–220.

181 Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. Biometrika. 1982;69:331–342.

182 Buzas JS. Unbiased scores in proportional hazards regression with covariate measurement error. Journal of Statistical Planning and Inference. 1998;p. 247–257.

183 Yi GY, Lawless JF. A corrected likelihood method for the proportional hazards model with covariates subject to measurement error. Journal of Statistical Planning and Inference. 2007;p. 1816–1828.

184 Keogh RH, Strawbridge AD, White IR. Effects of classical exposure measurement error on the shaoe of exposure-disease associations. Epidemiological Methods. 2012;1(1).

185 Breslow NE, Day NE, et al. Statistical methods in cancer research. vol. 1. International Agency for Research on Cancer Lyon; 1980.

186 Richardson DB, Loomis D. The impact of exposure categorisation for grouped analyses of cohort data. Occupational and Environmental Medicine. 2004;61(11):930–5.

187 Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. American Journal of Epidemiology. 1991;134(10):1233–1246.

188 Keogh RH, Strawbridge AD, White IR. Correcting for bias due to misclassification when error prone continuous exposures are misclassified. Epidemiological Methods. 2012;1(1).

189 Hertz-Picciotto I, Smith AH. Observations on the dose-response curve for arsenic exposure and lung cancer. Scandinavian Journal of Work, Environment & Health. 1993;19:217–226.

190 Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. International Journal of Epidemiology. 2006;35(4):1074–81.

191 Muff S, Riebler A, Held L, Rue H, Saner P. Bayesian analysis of measurement error models using integrated nested laplace approximation. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2015;64(2):231–252.

192 und Regina Schwarz TA. Cox's proportional hazars model under covariate measurement error - A review and comparison of methods. Sonderforschungsbereich. 2001;.

193 Richardson S, Gilks WR. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. American Journal of Epidemiology. 1993;138(6):430–442.

194 Richardson S, Gilks WR. Conditional independence models for epidemiological studies with covariate measurement error. Statistics in Medicine. 1993;12:1703–1722.

195 Congdon P. Bayesian Statistical Modelling. John Wiley & Sons; 2006.

196 Schafer DW, Purdy KG. Likelihood analysis for errors-in-variables regression with replicate measurements. Biometrika. 1996;83(4):813–824.

197 Küchenhoff H, Carroll RJ. Segmented regression with errors in predictors: semi-parametric and parametric methods. Statistics in Medicine. 1997;16:169–188.

198 Carroll RJ, Ruppert D, Crainiceanu CM, Tosteson TD, Karagas MR. Nonlinear and nonparametric regression and instrumental variables. Journal of the American Statistical Association. 2004;99(467):736–750.

199 Messer K, Natarajan L. Maximum likelihood, multiple imputation and regression calibration for measurement error adjustment. Statistics in Medicine. 2008;27(30):6332–50.

200 White IR. Commentary: dealing with measurement error: multiple imputation or regression calibration? International Journal of Epidemiology. 2006;35:1081–1082.

201 Sánchez BN, Kim S, Sammel MD. Estimators for longitudinal latent exposure models: examining measurement model assumptions. Statistics in Medicine. 2017;36(13):2048–2066.

202 Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research. 1981;18(1):39–50.

203 Bakker R. Re-measuring left-right: A comparison of SEM and Bayesian measurement models for extracting left-right party placements. Electoral Studies. 2009;28:413–421.

204 Goldsmith K, Chalder T, White PD, Sharpe M, Pickles A. Measurement error, time lag, unmeasured confounding: Considerations for longitudinal estimation of the effects of a mediator in randomised clinical trials. Statistical Methods in Medical Research. 2016;.

205 Murad H, Kipnis V, Freedman LS. Estimating and testing interactions when explanatory variables are subject to non-classical measurement error. Statistical Methods in Medical Research. 2016;25(5):1991–2013.

206 Wang L C Y adn Hsu, Feng ZD, Prentice RL. Regression calibration in failure time regression. Biometrics. 1997;53(131–145).

207 Liao X, Zucker DM, Li Y, Spiegelman D. Survival analysis with error-prone time-varying covariates: A risk set calibration approach. Biometrics. 2011;67(1):50–58.

208 Carroll RJ, Küchenhoff H, Lombard F, Stefanski LA. Asymptotics for the SIMEX estimator in nonlinear measurement error models. Journal of the American Statistical Association. 1996;91(433).

209 Spiegelman D, Logan R, Grove D. Regression calibration with heteroscedastic error variance. The International Journal of Biostatistics. 2011;7(1).

210 Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association. 1994;89(428):1314–1328.

211 Misumi M, Furukawa K, Cologne JB, Cullings HM. Simulation-extrapolation for bias correction with exposure uncertainty in radiation risk analysis utilizing grouped data. Journal of the Royal Statistical Society: Series C (Applied Statistics). 2017;.

212 Higdon R, Schafer DW. Maximum likelihood computations for regression with measurement error. Computational Statistics & Data Analysis. 2001;35:283 – 299.

213 Wolpert DH. The lack of a priori distrinctions between learning algorithms. Neural Computation. 1996;8(7):1341–1390.

214 Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Transactions on Evolutionary Computation. 1997;1(1):67–82.

215 Li Y, Lin X. Functional inference in frailty measurement error models for clustered survival data using the SIMEX approach. Journal of the American Statistical Association. 2003;98(461).

216 Hu P, Tsiatis AA, Davidian M. Estimating the parameters in the Cox model when covariate variables are measured with error. Biometrics. 1998;54:1407 – 1419.

217 Guolo A, Brazzale AR. A simulation-based comparison of techniques to correct for measurement error in matched case-control studies. Statistics in Medicine. 2008;27:3755–3775.

218 Torabi M. Likelihood inference in generalized linear mixed measurement error models. Computational Statistics & Data Analysis. 2013;p. 549–557.

219 Noh M, Wu L, Lee Y. Hierarchical likelihood methods for nonlinear and generalized linear mixed models with missing data and measurement errors in covariates. Journal of Multivariate Analysis. 2012;p. 42–51.

220 Li Y, Lin X. Covariate measurement error in frailty models for clustered survival data. Biometrika. 2000;87(4):849–866.

221 Little MP, Kukush AG, Masiuk SV, Shkylar S, Carroll RJ, Lubin JH, et al. Impact of uncertainties in expsoure assessment on estimates of thyroid cancer risk among Ukranian children and adolescents exposed from the Chernobyl accident. PLoS one. 2014;9(1).

222 Stayner L, Vrijheid M, Cardis E, Stram DO, Deltour I, Gilbert SJ, et al. A Monte Carlo maximum likelihood method for estimating uncertainty arising from shared errors in exposures in epidemiological studies of nuclear workers. Radiation Research. 2007;168(6):757–763.

223 Good IJ. The Bayesian influence, or how to sweep subjectivism under the carpet. In: Foundations of probability theory, statistical inference, and statistical theories of science. Springer; 1976. p. 125–174.

224 McGrayne SB. The theory that would not die: how Bayes' rule cracked the enigma code, hunted down Russian submarines & emerged triumphant from two centuries of controversy. Yale University Press; 2011.

225 Gelman A, Robert C. Rejoinder: The anti-Bayesian moment and its passing. The American Statistician. 2013;67(1).

226 Marin JM, Robert C. Bayesian core: a practical approach to computational Bayesian statistics. Springer Science & Business Media; 2007.

227 Robert C, Casella G. A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data. Statistical Science. 2011;26(1):102–115.

228 Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. Statistics in Medicine. 2005;24:2401–2428.

229 Hamelryck T. An overview of Bayesian inference and graphical models. In: Bayesian methods in structural bioinformatics. Springer Berlin Heidelberg; 2012. p. 3–48.

230 O'Hagan A. Eliciting expert beliefs in substantial practical applications. Journal of the Royal Statistical Society Series D (The Statistician). 1998;47(1):21–35.

231 Robert CP, Casella G. Monte Carlo statistical methods - Second Edition. Springer; 2004.

232 Brooks SP. Bayesian computation: a statistical revolution. Philosophical Transactions of the Royal Society of London. 2003;361:2681–2697.

233 Andrieu C, Doucet A, Robert CP. Computational advances for and from Bayesian analysis. Statistical Science. 2004;19(1):118–127.

234 Albert I, Ancelet S, David O, Denis JB, Makowski D, Parent E, et al. Initiation à la statistique bayésienne - Base théoriques et applications en alimentation, environnement, épidémiologie et génétique. Edition Ellipses; 2015.

235 Rao CR. R. A. Fisher: The founder of modern statistics. Statistical Science. 1992;7(1):34–48.

236 Krishnan T. Fisher's contributions to statistics. Resonance. 1997;2(9):32–37.

237 Sterne JAC, Smith GD. Sifting the evidence - what's wrong with significance tests? Physical Therapy. 2001;81(8).

238 Spanos A. Where do statistical models come from? Revisiting the problem of specification. In: Lecture Notes-Monograph Series Vol. 49, Optimality: The second Erich L. Lehmann Symposium. vol. 49; 2006. p. 98–119.

239 Efron B. Why isn't everyone a Bayesian. The American Statistician. 1986;40(1):1–5.

240 Loredo TJ. From Laplace to supernova SN 1987A: Bayesian inference in astrophysics. In: Maximum entropy and Bayesian methods. Springer; 1990. p. 81–142.

241 Fienberg SE. A brief history of statistics in three and one-half chapters: A review essay. Statistical Science. 1992;7(2):208–225.

242 Efron B. Bayesians, frequentists and scientists. Journal of the American Statistical Association. 2005;100(469):1–5.

243 Fienberg SE. When did Bayesian inference become "Bayesian". Bayesian Analysis. 2006;1(1):1–40.

244 Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. International Journal of Epidemiology. 2006;35:765–775.

245 Berger J. The case of objective Bayesian analysis. Bayesian Analysis. 2006;1(3):385–402.

246 Hald A. A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935. Springer Science & Business Media; 2008.

247 Efron B. The future of indirect evidence. Statistical Science. 2010;25(2).

248 Fienberg SE. Bayesian models and methods in public policy and government settings. Statistical Science. 2011;26(2).

249 Stigler SM. Thomas Bayes's Bayesian inference. Journal of the Royal Statistical Society: Series A (Statistics in Society). 1982;145:250–258.

250 Stigler SM. Laplace's 1774 memoir on inverse probability. Statistical Science. 1986;1(3):359–378.

251 Lindley DV. Understanding uncertainty. John Wiley & Sons; 2006.

252 Bayes T. An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London. 1763;53:370–418.

253 Bijak J, Bryant J. Bayesian demography 250 years after Bayes. Population Studies. 2016;70(1):1–19.

254 Stigler SM. Stigler's law of eponymy. Transactions of the New York Academy of Sciences. 1980;39(1 Series II):147–158.

255 Merton RK. Priorities in scientific discovery: A chapter in the sociology of science. American Sociological Review. 1957;22(6).

256 Block DL. Georges Lemaître and Stigler's law of eponymy. In: Georges Lemaître: Life, science and legacy. Springer Berlin Heidelberg; 2013. .

257 Stigler SM. Poisson on the Poisson distribution. Statistics & Probability Letters. 1982;1(1):33–35.

258 Gaddum JH. Lognormal distributions. Nature. 1945;156:463–466.

259 Le Cam L, Yo Lang G. Asymptotics in Statistics: Some Basic Concepts (second ed.). Springer; 2000.

260 Goodman S. A comment on replication, p-values and evidence. Statistics in Medicine. 1992;11:875–879.

261 Lehmann EL. The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two? Journal of the American Statistical Association. 1993;88(424):1242–1249.

262 Kadane JB. Prime time for Bayes. Controlled Clinical Trials. 1995;16:313–318.

263 Efron B. R. A. Fisher in the 21st century. Statistical Science. 1998;13(2):95–122.

264 Gigerenzer G. Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In: Wright G, Ayton P, editors. Subjective probability. Wiley; 1994. p. 129–161.

265 Hubbard R, Bayarri MJ. Confusion over measures of evidence (p's) versus errors (α's) in classical statistical testing. The American Statistician. 2003;57(3).

266 Hubbard R, Lindsay RM. Why p values are not a useful measure of evidence in statistical significance testing. Theory & Psychology. 2008;18(1):69–88.

267 Berger JO. Could Fisher, Jeffreys and Neyman have agreed on testing? Statistical Science. 2003;18(1):1–32.

268 Gelman A, Hennig C. Beyond subjective and objective in statistics. arXiv preprint arXiv:150805453. 2015;.

269 Zabell SL. R A Fisher and the fiducial argument. Statistical Science. 1992;7(3):369–387.

270 Lenhard J. Models and statistical inference: The controversy between Fisher and Neyman-Pearson. British Journal of the Philosophy of Science. 2006;57:69–91.

271 Lehmann EL. Fisher, Neyman, and the creation of classical statistics. Springer Science & Business Media; 2011.

272 Aldrich J. R. A. Fisher on Bayes and Bayes' theorem. Bayesian Analysis. 2008;3(1):161–170.

273 Fisher R. Statistical methods and scientific induction. Journal of the Royal Statistical Society: Series B (Methodological). 1955;17(1):69–78.

274 Box GEP. An apology for ecumensim in statistics. Wisconsin University - Madison mathematics research center; 1982.

275 Zabell S. R A Fisher on the history of inverse probability. Statistical Science. 1989;4(3):247–263.

276 Fienberg SE. Comment: Bayesian ideas reemerged in the 1950s. The American Statistician. 2013;67(1).

277 Berger JO. Bayesian analysis: a look at today and thoughts of tomorrow. Journal of the American Statistical Association. 2000;95(452):1269–1276.

278 Simpson D, Rue H, Martins TG, Riebler A, Sørbye SH. Penalising model component complexity: A principled, practical appraoch to constructing priors. Statistical Science. 2017;32(1):1–28.

279 O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, et al. Uncertain judgements: eliciting experts' probabilities. John Wiley & Sons; 2006.

280 G S. The unity and diversity of probability. Statistical Science. 1990;5(4):435–444.

281 Wassermann L. All of statistics: A concise course in statistical inference. Springer Science & Business Media; 2013.

282 Ore O. Pascal and the invention of probability theory. The American Mathematical Monthly. 1960;67(5):409–419.

283 Shafer G. The early development of mathematical probability. Companion Encyclopedia of the history and philosophy of the mathematical sciences. 1993;2:1293–1302.

284 Hacking I. The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference. Cambridge university press; 2006.

285 O'Hagan T. Dicing with the unknown. Significance. 2004;1(3):132–133.

286 Diaconis P, Mazur BC. The problem of thinking too much. Bulletin of the American Academy of Arts and Sciences. 2003;56(3):26–38.

287 Breuer T. The impossibility of accurate state self-measurements. Philosophy of Science. 1995;62(2):197–214.

288 Binder PM. Philosophy of science: Theorie of almost everything. Nature. 2008;455(7215):884–885.

289 Nau RF. De Finetti was right: Probability does not exist. Theory and Decision. 2001;51:89–124.

290 Galavotti MC. Subjectivism, objectivism and objectivity in Bruno de Finetti's Bayesianism. In: Corfield D, Williamson J, editors. Foundations of Bayesianism. Springer; 2001. .

291 Gelfand AE, Smith AF. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association. 1990;85(410):398–409.

292 Robert C. The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media; 2007.

293 Gelman A, Shalizi CR. Philosophy and practice of Bayesian statistics. British Journal of Mathematical and Statistical Psychology. 2013;66(1):8–38.

294 Robert CP. The expected demise of the Bayes factor. Journal of Mathematical Psychology. 2016;72:33–37.

295 Dennis B. Discussion: should ecologists become Bayesians? Ecological Applications. 1996;6(4):1095–1103.

296 Savage LJ. The foundations of statistics reconsidered. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. The Regents of the University of California; 1961. .

297 Savage LJ. On rereading R. A. Fisher. The Annals of Statistics. 1976;1(3):441–500.

298 Green PJ, Łatuszyński K, Pereyra M, Robert CP. Bayesian computation: a summary of the current state, and samples backwards and forwards. Statistics and Computing. 2015;25(4):835–862.

299 Gilks WR, Richardson S, Spiegelhalter DJ. Markov chain Monte Carlo in practice. Boca Raton: Chapman Hall; 1996.

300 Roberts GO, Rosenthal JS. Examples of adaptive MCMC. Journal of Computational and Graphical Statistics. 2009;18(2):349–367.

301 Besag J, Green PJ. Spatial statistics and Bayesian computation. Journal of the Royal Statistical Society: Series B (Methodological). 1993;55(1):25–37.

302 Metropolis N. The beginning of the Monte Carlo method. Los Alamos Science Special Issue. 1987;p. 125–130.

303 Landau DP, Binder K. A guide to Monte-Carlo simulations in statistical physics. Cambridge university press; 2014.

304 Shonkwiler RW, Mendivil F. Explorations in Monte Carlo methods. Springer Science & Business Media; 2009.

305 Roberts GO, Smith AF. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. Stochastic processes and their applications. 1994;49(2):207–216.

306 Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992;p. 457–472.

307 Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. The Journal of Chemical Physics. 1953;21(6):1087–1092.

308 Gubernatis JE. Marshall Rosenbluth and the Metropolis algorithm. Physics of Plasmas. 2005;12(5).

309 Anderson HL. Scientific uses of the MANIAC. Journal of Statistical Physics. 1986;43(5–6):731–748.

310 Holian B. A history of consitutive modeling via molecular dynamics: Shock waves in fluids and gases. In: EPJ Web of Conferences. vol. 10; 2010. .

311 Waltman L. An empirical analysis of the use of alphabetical authorship in scientific publishing. Journal of Informetrics. 2012;6(4):700–711.

312 Hastings WK. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika. 1970;57(1):97–109.

313 Andrieu C, Doucet A, Holenstein R. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2010;72:269–342.

314 Chib S, Greenberg E. Understanding the metropolis-hastings algorithm. The American Statistician. 1995;49(4):327–335.

315 Morris WK, Vesk PA, McCarthy MA. Profiting from pilot stides: Analysing mortality using Bayesian models with informative priors. Basic and Applied Ecology. 2012;.

316 Geneletti S, O'Keefe AG, Sharples LD, Richardson S, Baio G. Bayesian regression discontinuity designs: Incorporating clinical knowledge in the causal analysis of primary care data. Statistics in Medicine. 2015;34(15):2334–2352.

317 Jaynes ET. Highly informative priors. In: Bernardo JM, M H DeGroot DVL, Smith AFM, editors. Bayesian statistics 2; 1985. p. 329–360.

318 Parent E, Bernier J. Encoding prior experts judgements to improve risk analysis of extreme hydrological events via POT modeling. Journal of Hydrology. 2003;283:1–18.

319 Goldstein M. Subjective Bayesian analysis: Principles and practice. Bayesian Analysis. 2006;1(3):403–420.

320 Greenland S. Bayesian perspectives for epidemiological research: II. Regression analysis. International Journal of Epidemiology. 2007;36:195–202.

321 Miranda-Moreno LF, Heydari S, Lord D, Fu L. Bayesian road safety analysis: Incorporation of past evidence and effect of hyper-prior choice. Journal of Safety Research. 2013;46:31–40.

322 Gustafson P. Robustness considerations in Bayesian analysis. Statistical Methods in Medical Research. 1996;5:357–373.

323 Kass RE, Wassermann L. The selection of prior distributions by formal rules. Journal of the American Statistical Association. 1996;91(435):1343–1370.

324 Bernardo JM. Reference posterior distirbutions for Bayesian inference. Journal of the Royal Statistical Society: Series B (Methodological). 1979;41(2):113–147.

325 Guo X, Carlin BP. Seperate and joint modeling of longitudinal and event time data using standard computer packages. The American Statistician. 2004;58(1).

326 Gelman A. The boxer, the wrestler, and the coin flip: a paradox of robust Bayesian inference and belief functions. The American Statistician. 2006;60(2):146–150.

327 de Finetti B. Bayesianism: Its unifying role for both the foundations and applications of statistics. International Statistical Review. 1974;42(2):117–130.

328 Berger JO, Bernardo JM. On the development of reference priors. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. Bayesian statistics 4. Oxford University Press; 1992. p. 35–60.

329 Berger JO, Bernardo JM, Sun D. The formal definition of reference priors. The Annals of Statistics. 2009;37(2):905–938.

330 Gelman A. Bayes, Jeffreys, prior distributions and the philosophy of statistics. Statistical Science. 2009;24(2):176–178.

331 Parent E, Favre AC, Bernier J, Perreault L. Copula models for frequency analysis what can be learned from a Bayesian perspective. Advances in Water Resources. 2014;63(91–103).

332 Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. vol. 2. Taylor & Francis; 2014.

333 Held L, Bové DS. Applied statistical inference - Likelihood and Bayes. Springer; 2014.

334 Robert CP, Chopin N, Rousseau J. Harold Jeffreys's theory of probability revisited. Statistical Science. 2009;24(2):141–172.

335 Bernardo JM. Reference analysis. In: Handbook of statistics. Elsevier; 2005. p. 17–90.

336 Gelman A. Bayesian model-building by pure thought: some principles and examples. Statistica Sinica. 1996;6:215–232.

337 Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis. 2006;1(3):515–533.

338 Burke DL, Bujkiewicz S, Riley RD. Bayesian bivariate meta-analysis of correlated effects: Impact of the prior distributions on the between-study correlation, borrowing of strength, and joint inferences. Statistical Methods in Medical Research. 2016;.

339 Gelman A. P values and statistical practice. Epidemiology. 2013;24(1):69–72.

340 Bayarri MJ, Berger JO. The interplay of Bayesian and frequentist analysis. Statistical Science. 2004;19(1):58–80.

341 Lindley DV. The use of prior probability distributions in statistical inference and decision. In: Proc. 4th Berkeley Symp. on Math. Stat. and Prob; 1961. p. 453–468.

342 Fienberg SE. Does it make sense to be an "objective Bayesian"? (Comment on articles by Berger and Goldstein). Bayesian Analysis. 2006;1(3):429–432.

343 Kadane JB. Comment on article by Gelman. Bayesian Analysis. 2008;3(3):455–458.

344 Albert I, Donnet S, Guihenneuc-Jouyaux C, Low-Choy S, Mengersen K, Rousseau J, et al. Combining expert opinions in prior elicitation. Bayesian Analysis. 2012;7(3):503–532.

345 Senn S. You may believe you are a Bayesian but you are probably wrong. Rationality, Markets and Morals. 2011;2(42):48–66.

346 Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. Risk Analysis. 1999;19(2):187–203.

347 Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. Journal of the American Statistical Association. 2005;100(470):680–701.

348 Zapata-Vásquez RE, O'Hagan A, Bastos LS. Eliciting expert judgements about a set of proportions. Journal of Applied Statistics. 2014;.

349 Cooke RM, Goossens LL. TU Delft expert judgment data base. Reliability Engineering & System Safety. 2008;93(5):657–674.

350 Soares MO, Bojke L, Dumville J, Iglesias C, Cullum N, Claxton K. Methods to elicit experts' beliefs over uncertain quantities: application to a cost effectiveness transition model of negative pressure wound therapy for severe pressure ulceration. Statistics in Medicine. 2011;30(19):2363–2380.

351 Budescu D, Abbas A, Wu L. Does probability weighting matter in probability elicitation. Journal of Mathematical Psychology. 2011;55:320–327.

352 Cox R, Sanchez J, Revie CW. Multi-criteria decision analysis tools for prioritising emerging or re-emerging infectious diseases associated with climate change in Canada. PLoS One. 2013;8(8):e68338.

353 Edlmann K, Bensabat J, Niemi A, Haszeldine RS, McDermott CI. Lessons learned from using expert elicitation to identify, assess and rank the potential leakage scenarios at the Heletz pilot CO2 injection site. International Journal of Greenhouse Gas Control. 2016;49:473–487.

354 World Health Organization, Foodborne Epidemiology Reference Group, Source Attribution Task Force. Research synthesis methods in an age of globalized risks: Lessons from the global burden of foodborne disease expert elicitation. Risk Analysis. 2016;36(2):191–202.

355 Tversky A, Kahneman D. Availability: a heuristic for judging frequency and probability. Cognitive Psychology. 1973;5:207–232.

356 Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1974;185:1124–1131.

357 Genest C, Zidek JV. Combining probability distributions: A critique and annotated bibliography. Statistical Science. 1986;1(1):114–148.

358 Kadane JB, Wolfson LJ. Experiences in elicitation. Journal of the Royal Statistical Society Series D (The Statistician). 1998;47:3–19.

359 Baddeley MC, Curtis A, Wood R. An introduction to prior information derived from probabilistic judgements: elicitation of knowledge, cogntive bias and herding. Geological Society, London. 2004;239(1):15–27.

360 Kynn M. The 'heuristics and biases' bias in expert elicitation. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2008;171:239–264.

361 Wolfson LJ, Bousquet N. Elicitation. In: Wiley StatsRef: Statistics Reference online. John Wiley & Sons; 2016. p. 1–11.

362 van Dorp JR, Mazzuchi TA. Statistics Textbooks and Monographs. Statistics Textbooks and Monographs. 2004;174:283–318.

363 Cook JD. Determining distribution parameters from quantiles. Department of Biostatistics of the University of Texas; 2010.

364 Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. Psychological Review. 1995;102(4):684–704.

365 Gigerenzer G. On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). Psychological Review. 1996;103(3):592–596.

366 Cosmides L, Tooby J. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judment under uncertainty. Cognition. 1996;58:1–73.

367 Sedlmeier P, Gigerenzer G. Teaching Bayesian reasoning in less than two hours. Journal of Experimental Psychology: General. 2001;130(3):380–400.

368 Low Choy S, James A, Mengerson K. Expert elicitation and its interface with technology: a review with a view to designing Elicitator. In: 18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009; 2009. .

369 Burgman MA, McBride M, Ashton R, Speirs-Bridge A, Flander L, Wintle B, et al. Expert status and performance. PLoS One. 2011;6(7):e22998.

370 Ferrari P, Carroll RJ, Gustafson P, Riboli E. A Bayesian multilevel model for estimating the diet/disease relationship in a multicenter study with exposures measured with error: The EPIC study. Statistics in Medicine. 2008;27:6037–6054.

371 Greenland S. Principles of multilevel modelling. International Journal of Epidemiology. 2000;29(158–167).

372 Richardson S. Measurement error. In: Gilks WR, Richardson S, Spiegelhalter DJ, editors. Markov chain Monte Carlo in practice. Chapman & Hall; 1996. p. 401–417.

373 Richardson S, Leblond L, Jaussent I, Green PJ. Mixture models in measurement error problems, with reference to epidemiological studies. Journal of the Royal Statistical Society: Series A (Statistics in Society). 2002;165(3):549–566.

374 Clayton DG. Models for the analysis of cohort and case-control studies with inaccurately measured exposures. In: Dwyer JH, Feinleib M, Hoffmeister H, editors. Statistical models for longitudinal studies of health. Oxford University Press; 1992. p. 301–331.

375 Jordan MI. Graphical models. Statistical Science. 2004;19(1):140–155.

376 Mallick BK, Gelfand AE. Semiparametric errors-in-variables models - A Bayesian approach. Journal of Statistical Planning and Inference. 1996;52:307–321.

377 Berry SM, Carroll RJ, Ruppert D. Bayesian smoothing and regression splines for measurement error problems. Journal of the American Statistical Association. 2002;97(457):160–169.

378 Lambert P, Eilers PHC. Bayesian proportional hazards model with time-varying regression coefficients: A penalized Poisson regression approach. Statistics in Medicine. 2005;24:3977–3989.

379 Laird N, Olivier D. Covariance analysis of censored survival data using log-linear analysis techniques. Journal of the American Statistical Association. 1981;76(374):231–240.

380 Ibrahim JG, Chen MH, Sinha D. Bayesian survival analysis. New York: Springer; 2001.

381 Kleinbaum DG, Klein M. Survival analysis. New York: Springer; 1996.

382 Klein JP, Moeschberger MM. Survival analysis - Techniques for censored and truncated data. Springer; 2003.

383 Fahrmeir L, Hennerfeind A. Nonparametric Bayesian hazard rate models based on penalized splines. Discussion paper Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München; 2003.

384 Hauptmann M, Berhane K, Langholz B, Lubin J. Using splines to analyse latency in the Colorado Plateau uranium miners cohort. Journal of Epidemiology and Biostatistics. 2001;6(6):417–424.

385 Langholz B, Thomas D, Xiang A, Stram D. Latency analysis in epidemiologic studies of occupational exposures: application to the Colorado Plateau uranium miners cohort. American Journal of Industrial Medicine. 1999;35(3):246–256.

386 Richardson DB, Cole SR, Chu H, Langholz B. Lagging exposure information in cumulative exposure-response analyses. American Journal of Epidemiology. 2011;p. 260.

387 Thiébaut A, Bénichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. Statistics in Medicine. 2004;23(24):3803–3820.

388 Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. American Journal of Epidemiology. 1997;145(1):72–80.

389 Cologne J, Hsu WL, Abbott RD, Ohishi W, Grant EJ, Fujiwara S, et al. Proportional hazards regression in epidemiologic follow-up studies: an intuitive consideration of primary time scale. Epidemiology. 2012;23(4):565–573.

390 Wolkewitz M, Allignol A, Harbarth S, de Angelis G, Schumacher M, Beyersmann J. Time-dependent study entries and exposures in cohort studies can easily be sources of different and avoidable types of bias. Journal of Clinical Epidemiology. 2012;65(11):1171–80.

391 Therneau T, Grambsch P. Modeling Survival Data, Extending the Cox model. Springer; 2000.

392 Therneau T, Crowson C. Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model. The survival Package (R help guide). 2013;.

393 van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. Journal of Clinical Epidemiology. 2004;57(7):672–82.

394 Beyersmann J, Gastmeier P, Wolkewitz M, Schumacher M. An easy mathematical proof showed that time-dependent bias inevitably leads to biased effect estimation. Journal of Clinical Epidemiology. 2008;61(12):1216–21.

395 Barnett AG, Beyersmann J, Allignol A, Rosenthal VD, Graves N, Wolkewitz M. The time-dependent bias and its effect on extra length of stay due to nosocomial infection. Value Health. 2011;14(2):381–6.

396 Martino S, Akerkar R, Rue H. Approximate Bayesian inference for survival models. Scandinavian Journal of Statistics. 2011;38(3).

397 Tadesse MG, Ibrahim JG, Gentleman R, Chiaretti S, Ritz J, Foa R. Bayesian Error-in-Variable Survival Model for the Analysis of GeneChip Arrays. Biometrics. 2005;61(2):488–497.

398 Li. Bayesian proportional hazard analysis of the timing of high school dropout decisions. Econometric Reviews. 2007;26(5):529–556.

399 He B, Luo S. Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinson's disease. Statistical Methods in Medical Research. 2014;.

400 Caër-Lorho S. Base de données des taux de mortalité de référence, Rapport interne, IRSN/PRP-HOM/SRBE/LEPID/2013-07. Institut de Radioprotection et de Sureté Nucléaire; 2013.

401 Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiology Biomarkers & Prevention. 2010;19(8):1893–1907.

402 Heid I. Measurement error in exposure assessment: an error model and its impact on studies on lung cancer and residential radon exposure in Germany (thesis). Ludwig-Maximilians-Universität; 2002.

403 Hunter N, Muirhead CR, Miles JCH. Two error components model for measurement error: application to radon in homes. Journal of Environmental Radioactivity. 2011;102:799–805.

404 el Hussein A, Mohammed A, El-Hady MA, Ahmed AA, Ali AE, Barakat A. Diurnal and seasonal variation of short-lived radon progeny concentration and atmospheric temporal variations of 210Ph and 7Be in Egypt. Atmospheric Environment. 2001;35:4305–4313.

405 Flüry-Herard EBA, Paquet F. Les méthodes et les limites de la dosimétrie après contamination interne. Radioprotection. 2007;42(4):501 – 517.

406 Kreuzer M, Grosche B, Dufey F, Schnelzer M, Tschense A, Walsh L. The German uranium miners cohort study (Wismut cohort), 1946 - 2003. Bundesamt für Strahlenschutz; 2011.

407 Porstendörfer J, Reineking A. Radon: Characteristics in air and dose conversion factors. Health Physics. 1999;76(3):300–305.

408 Bouland D, Chouard JC. Submicron-sized aerosol and radon progeny measurements in a uranium mine. Radiation Protection Dosimetry. 1992;45(1–4):91–94.

409 Bigu J. Electrical charge characteristics of long-lived radioactive dust. Health Physics. 1990;58(3):341–350.

410 Porstendörfer J. Physical parameters and dose factors of the radon and thoron decay products. Radiation Protection Dosimetry. 2001;94(4):365–373.

411 Marsh JW, Birchall A. Uncertainty analyis of the absorbed dose to regions of the lung per unit exposure to radon progeny in a mine. Health Protection Agency, Radiation Protection Division, Chilton, Didcot, UK; 2008.

412 Ruzer LS, V NA, Harley NH. Assessment of lung deposition and breathing rate of underground miners in Tadjikistan. Radiation Protection Dosimetry. 1995;58:261–268.

413 DasGupta A. Probability for statistics and machine learning - Fundamentals and advanced topics. Springer Science + Business Media; 2011.

414 French S. Comment on Article by Albert et al. Bayesian Analysis. 2012;7(3):533–536.

415 Rue H, Martino S, Chopin N. Approximate Bayesian Inference for latent Gaussian models by using integrated nested laplace approximation. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2009;71(2).

416 Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK. Bayesian computing with INLA: A review. Annual Review of Statistics and its Application. 2017;4:395–421.

417 Höppner F, Klawonn F. Clustering with size constraints. In: Jain LC, Sato-Ilic M, Virvou M, Tsihrintzis GA, Balas VE, Abeynayake C, editors. Computational Intelligence Paradigms. Springer; 2008. p. 167–180.

418 Zhu S, Wang D, Li T. Data clustering with size constraints. Knowledge-based systems. 2010;23(8):883–889.

419 Dellaportas P, Roberts GO. An introduction to MCMC. In: Spatial statistics and computational methods. Springer; 2003. .

420 Janzen D, Saiedian H. Test-driven development concepts, taxonomy, and future direction. Computer. 2005;38(9):43–50.

421 Erdogmus H, Melnik G, Jeffries R. Test-driven development. In: Laplante PA, editor. Encyclopedia of Software Engineering; 2011. p. 1211–1229.

422 Maximilien EM, Williams L. Assessing test-driven development at IBM. In: Proceedings - International Conference on Software Engineering. vol. 6; 2003. p. 564–569.

423 Williams L, Maximilien EM, Vouk M. Test-driven development as a defect-reduction practice. In: 14th International Symposium on Software Reliability Engineering.; 2003. p. 34–45.

424 Bhat T, Nagappan N. Evaluating the efficacy of test-driven development: industrial case studies. In: Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering; 2006. p. 356–363.

425 Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2002;64(4):583–639.

426 Spiegelhalter D, Best NG, Carlin BP, van der Linde A. The deviance information criterion: 12 years on. Journal of the Royal Statistical Society: Series B (Methodological). 2014;76(3):485–493.

427 Land CE, Kwon D, Hoffman FO, Moroz B, Drozdovitch V, Bouville A, et al. Accounting for Shared and Unshared Dosimetric Uncertainties in the Dose Response for Ultrasound-Detected Thyroid Nodules after Exposure to Radioactive Fallout. Radiation Research. 2015;183(2):159–173.

428 Sylvestre MP, Abrahamowicz M. Comparison of algorithms to generate event times conditional on time-dependent covariates. Statistics in Medicine. 2008;27(14):2618–34.

429 Austin PC. Generating survival times to simulate Cox proportional hazards models with time-varying covariates. Statistics in Medicine. 2012;31:3946 – 3958.

430 Hendry DJ. Data generation for the Cox proportional hazards model with time-dependent covariates: a method for medical researchers. Statistics in Medicine. 2014;33:436–454.

431 Austin PC. A comparison of regression trees, logistic regression, generalized additive models, and multivariate adaptive regression splines for predicting AMI mortality. Statistics in Medicine. 2007;26(15):2937–57.

432 Zhou M. Understanding the Cox regression model with time-change covariates. The American Statistician. 2001;55(2):153–155.

433 Steenland K, Deddens J, Zhao S. Biases in estimating the effect of cumulative exposure in log-linear models when estimated exposure levels are assigned. Scandinavian Journal of Work, Environment & Health. 2000;26(1):37–43.

434 Stefanski LA. Measurement error models. In: Raftery AE, Tanner MA, Wells MT, editors. Statistics in the 21st Century. CHAPMAN and HALL; 2002. .

435 Van Roesbroeck S, Ruifeng L, Hoek G, Lebret E, Brunekreef B, Spiegelman D. Traffic-related outdoor pollution and respiratory symptoms in children - The impact of adjustment for exposure measurement error. Epidemiology. 2008;19(3):409–416.

436 Gössl C, Küchenhoff H. Bayesian analysis of logistic regression with an unknown change point and covariate measurement error. Statistics in Medicine. 2001;20:3109–3121.

437 Hoffman MD, Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research. 2014;15(1):1593–1623.

438 Blair A, Stewart PA, Zaebst DD, Pottern L, Zey JN, Bloom TF, et al. Mortality of industrial workers exposed to acrylonitrile. Scandinavian Journal of Work, Environment & Health. 1998;24 Suppl 2:25–41.

439 Thiébaut ACM, Kipnis V, Chang SC, Subar AF, Thompson FE, Rosenberg PS, et al. Dietary fat and postmenopausal invasive breast cancer in the National Institutes of Health-AARP Diet and Health Study cohort. Journal of the National Cancer Institute. 2007;99(6):451–62.

440 Hart JE, Liao X, Hong B, Puett RC, Yanosky JD, Suh H, et al. The association of long-term exposure to PM2.5 on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. Environ Health. 2015;14:38.

441 Khan AH, Puranik VD. Radiation protection and environmental safety surveillance in uranium mining and ore processing in India. In: Merkel B, Schipek M, editors. The new uranium mining boom. Springer Berlin Heidelberg; 2012. .

442 Banzi FP, Msaki P, Mohammed N. Challenging issues in regulating uranium mining in Tanzania. In: Merkel BJ, Arab A, editors. Uranium - past and future challenges. Springer International Publishing Switzerland; 2015. .

443 Falck WE, Coetzee H. Making uranium-mining more sustainable - the FP7 project EO-MINERS. In: Merkel B, Schipek M, editors. The new uranium mining boom. Springer; 2011. .

444 Banzi FP, Msaki P, Mohammed N. Uranium boom in Namibia - Hausse or Baisse. In: Merkel BJ, Arab A, editors. Uranium - past and future challenges. Springer International Publishing Switzerland; 2015. .

445 Darby SC, Hill DC. Health effects of residential radon: a European perspective at the end of 2002. Radiation Protection Dosimetry. 2003;104(4):321–329.

446 Taeger D, Krahn U, Wiethege T, Ickstadt K, Johnen G, Eisenmenger A, et al. A study on lung cancer mortality related to radon, quartz, and arsenic exposures in German uranium miners. Journal of Toxicology and Environmental Health, Part A. 2008;71(13-14):859–865.

447 Sun Y, Bochmann F, Nold A, Mattenklott M. Diesel Exhaust Exposure and the Risk of Lung Cancer—A Review of the Epidemiological Evidence. International Journal of Environmental Research and Public Health. 2014;11(2):1312–1340.

448 Brüske-Hohlfeld I, Rosario AS, Wölke G, Heinrich J, Kreuzer M, Kreienbrock L, et al. Lung cancer risk among former uranium miners of the WISMUT Company in Germany. Health Physics. 2006;90(3):208–216.

449 Walsh L, Dufey F, Möhner M, Schnelzer M, Tschense A, Kreuzer M. Differences in baseline lung cancer mortality between the German uranium miners cohort and the population

of the former German Democratic Republic (1960-2003). Radiation and Environmental Biophysics. 2011;50(1):57–66.

450 Krewski D, Lubin JH, Zielinski JM, Alavanja M, Catalan VS, Field RW, et al. Residential radon and risk of lung cancer: a combined analysis of 7 North American case-control studies. Epidemiology. 2005;16(2):137–145.

451 Richardson DB, Laurier D, Schubauer-Berigan MK, Tchetgen Tchetgen E, Cole SR. Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models. American Journal of Epidemiology. 2014;180(9):933–40.

452 Keil AP, Richardson DB, Troester MA. Healthy worker survivor bias in the Colorado Plateau uranium miners cohort. American Journal of Epidemiology. 2015;181(10):762–770.

453 Puskin JS, James CA. Radon Exposure assessment and dosimetry applied to epidemiology and risk estimation. Radiation Research. 2006;166:193–208.

454 Chambers DB, Stager RH. Prediction of the variation in risks from exposure to radon at home or at work. Radiation Protection Dosimetry. 2011;146(1-3):34–7.

455 Susser M. The logic of Sir Karl Popper and the practice of epidemiology. American Journal of Epidemiology. 1986;124.

456 Sims CA. Pitfalls of a minimax approach to model uncertainty. The American Economic Review. 2001;91(2):51–54.

Appendices

Appendix A

Accelerating the evaluation of the posterior distribution

A.1 Accelerating the evaluation of the measurement model

In order to speed up inference, we accelerated the evaluation of the measurement model by using the rules that additions are faster then multiplications and that applying the exponential function is very computer intensive. In particular, in the measurement model, we are concerned with the evaluation of $[X_{iq}|Z_{iq}, \boldsymbol{\sigma^B}]$ in the case of Berkson error and $[Z_{iq}|X_{iq}, \boldsymbol{\sigma^c}]$ in the case of classical measurement error. As described in section 5.1.2., we made the assumption that X_{iq} conditional on Z_{iq} in the Berkson model and Z_{iq} given X_{iq} in the classical measurement error model follow a lognormal distribution. The density of a variable Y that follows a lognormal distribution is given by

$$\frac{1}{y} \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right) \tag{A.1}$$

where μ and σ are the mean and the standard deviation of its natural logarithm, respectively. Its mean is given by $E(Y) = \exp(\mu + \frac{\sigma^2}{2})$. As we assumed Berkson and classical measurement error without a systematic bias component, we want the two measurement models to verify $E(X_{iq}|Z_{iq}) = Z_{iq}$ for Berkson error and $E(Z_{iq}|X_{iq}) = X_{iq}$ for classical measurement error. Accordingly, we chose the parameters of the conditional distribution of X_{iq} given Z_{iq} as $\mu = \ln(Z_{iq}) + \frac{\sigma^2_{p_{iq}}}{2}$ and $\sigma_{p_{iq}}$. For the conditional distribution of Z_{iq} given X_{iq} we chose a lognormal distribution $\mu = \ln(X_{iq}) + \frac{\sigma^2_{p_{iq}}}{2}$ and $\sigma_{p_{iq}}$.

Measurement model for Berkson error

Given these parameter choices, we can simplify the Metropolis-Hastings ratio that has to be evaluated to update the values of true exposure X_{iq} in the case of Berkson error by the following

simplifications:

$$\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{f\left(X_{iq}^{\text{cand}}|Z_{iq},\boldsymbol{\sigma}\right)}{f\left(X_{iq}^{\text{t}}|Z_{iq},\boldsymbol{\sigma}\right)}=\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\frac{1}{X_{iq}^{\text{cand}}\sigma_{p_{iq}}\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\text{cand}}\right)-\ln(Z_{iq})+\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}{\frac{1}{X_{iq}^{\text{t}}\sigma_{p_{iq}}\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\text{t}}\right)-\ln(Z_{iq})+\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}$$

$$=\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{X_{iq}^{\text{t}}\cdot\exp\left(-\frac{\ln(X_{iq}^{\text{cand}})^2-2\ln(X_{iq}^{\text{cand}})\left(\ln(Z_{iq})-\frac{\sigma_{p_{iq}}^2}{2}\right)+\left(\ln(Z_{iq})-\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}{X_{iq}^{\text{cand}}\cdot\exp\left(-\frac{\ln(X_{iq}^{\text{t}})^2-2\ln(X_{iq}^{\text{t}})\left(\ln(Z_{iq})-\frac{\sigma_{p_{iq}}^2}{2}\right)+\left(\ln(Z_{iq})-\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}$$

$$=\exp\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\frac{\ln(X_{iq}^{\text{t}})^2-\ln(X_{iq}^{\text{cand}})^2+\left(2\ln(Z_{iq})-3\sigma_{p_{iq}}^2\right)\ln\left(\frac{X_{iq}^{\text{cand}}}{X_{iq}^{\text{t}}}\right)}{2\sigma_{p_{iq}}^2}\right)$$

Measurement model for shared Berkson error

Similarly to unshared Berkson error, we can simplify the Metropolis-Hastings ratio for shared Berkson error in the following way:

$$
\begin{aligned}
\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{f\left(U_{iq}^{\text{cand}}|\boldsymbol{\sigma}\right)}{f\left(U_{iq}^{\text{t}}|\boldsymbol{\sigma}\right)} &= \prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\frac{1}{U_{iq}^{\text{cand}}\sigma_{p_{iq}}\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(U_{iq}^{\text{cand}}\right)+\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}{\frac{1}{U_{iq}^{\text{t}}\sigma_{p_{iq}}\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(U_{iq}^{\text{t}}\right)+\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)} \\
&= \prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{U_{iq}^{\text{t}}\cdot\exp\left(-\frac{\ln(U_{iq}^{\text{cand}})^2+2\ln(U_{iq}^{\text{cand}})\left(\frac{\sigma_{p_{iq}}^2}{2}\right)+\left(\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}{U_{iq}^{\text{cand}}\cdot\exp\left(-\frac{\ln(U_{iq}^{\text{t}})^2+2\ln(U_{iq}^{\text{t}})\left(\frac{\sigma_{p_{iq}}^2}{2}\right)+\left(\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)} \\
&= \prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\exp\left(-\frac{-2\ln(U_{iq}^{\text{t}})\sigma_{p_{iq}}^2+\ln(U_{iq}^{\text{cand}})^2+\ln(U_{iq}^{\text{cand}})\sigma_{p_{iq}}^2+\left(\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}{\exp\left(-\frac{-2\ln(U_{iq}^{\text{cand}})\sigma_{p_{iq}}^2+\ln(U_{iq}^{\text{t}})^2+\ln(U_{iq}^{\text{t}})\sigma_{p_{iq}}^2+\left(\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)} \\
&= \exp\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\frac{\ln(U_{iq}^{\text{t}})^2-\ln(U_{iq}^{\text{cand}})^2-3\sigma_{p_{iq}}^2\left(\ln(U_{iq}^{\text{cand}})-\ln(U_{iq}^{\text{t}})\right)}{2\sigma_{p_{iq}}^2}\right)
\end{aligned}
$$

Measurement model for classical measurement error

$$\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{f\left(Z_{iq}|X_{iq}^{\text{cand}},\boldsymbol{\sigma}\right)}{f\left(Z_{iq}|X_{iq}^{\text{t}},\boldsymbol{\sigma}\right)}=\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\frac{1}{Z_{iq}\sigma_{p_{iq}}\sqrt{2\pi}}\exp\left(-\frac{\left(\ln(Z_{iq})-\ln\left(X_{iq}^{\text{cand}}\right)+\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}{\frac{1}{Z_{iq}\sigma_{p_{iq}}\sqrt{2\pi}}\exp\left(-\frac{\left(\ln(Z_{iq})-\ln\left(X_{iq}^{\text{t}}\right)+\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}$$

$$=\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\exp\left(-\frac{\ln(X_{iq}^{\text{cand}})^2-2\ln(X_{iq}^{\text{cand}})\left(\ln(Z_{iq})+\frac{\sigma_{p_{iq}}^2}{2}\right)+\left(\ln(Z_{iq})+\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}{\exp\left(-\frac{\ln(X_{iq}^{\text{t}})^2-2\ln(X_{iq}^{\text{t}})\left(\ln(Z_{iq})+\frac{\sigma_{p_{iq}}^2}{2}\right)+\left(\ln(Z_{iq})+\frac{\sigma_{p_{iq}}^2}{2}\right)^2}{2\sigma_{p_{iq}}^2}\right)}$$

$$=\exp\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\frac{\ln(X_{iq}^{\text{t}})^2-\ln(X_{iq}^{\text{cand}})^2+\left(2\ln(Z_{iq})+\sigma_{p_{iq}}^2\right)\ln\left(\frac{X_{iq}^{\text{cand}}}{X_{iq}^{\text{t}}}\right)}{2\sigma_{p_{iq}}^2}\right)$$

Further simplification using the fact that $\sigma_{p_{iq}}^2=\sigma_5^2$ since classical error only occurred during the last calendar period (1983-1999) yields:

$$\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{f\left(Z_{iq}|X_{iq}^{\text{cand}},\boldsymbol{\sigma}\right)}{f\left(Z_{iq}|X_{iq}^{\text{t}},\boldsymbol{\sigma}\right)}=$$

$$\exp\left(\frac{1}{2\sigma_5^2}\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\left[\ln(X_{iq}^{\text{t}})^2-\ln(X_{iq}^{\text{cand}})^2+\left(2\ln(Z_{iq})+\sigma_5^2\right)\ln\left(\frac{X_{iq}^{\text{cand}}}{X_{iq}^{\text{t}}}\right)\right]\right)$$

A.2 Accelerating the evaluation of the exposure model

When updating the latent exposure values X, the evaluation of the exposure model can be simplified in order to accelerate inference:

$$\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\pi(X_{iq}^{\text{cand}})}{\pi(X_{iq}^{\text{t}})}=\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\frac{1}{X_{iq}^{\text{cand}}\sigma_X\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\text{cand}}\right)-\mu_X\right)^2}{2\sigma_X^2}\right)}{\frac{1}{X_{iq}^{\text{t}}\sigma_X\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\text{t}}\right)-\mu_X\right)^2}{2\sigma_X^2}\right)}$$

$$=\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{X_{iq}^{t}}{X_{iq}^{\text{cand}}}\exp\left(\frac{\ln(X_{iq}^{t})^2-\ln(X_{iq}^{\text{cand}})^2+2\ln\left(\frac{X_{iq}^{\text{cand}}}{X_{iq}^{\text{t}}}\right)\cdot\mu_X}{2\sigma_X^2}\right)$$

$$=\exp\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\frac{\ln(X_{iq}^{t})^2-\ln(X_{iq}^{\text{cand}})^2+2(\mu_X-\sigma_X^2)\ln\left(\frac{X_{iq}^{\text{cand}}}{X_{iq}^{\text{t}}}\right)}{2\sigma_X^2}\right)$$

Similarly, when using a Metropolis-Hastings step to update the parameters μ_x and σ_x we can simplify this ratio by writing

$$\begin{aligned}\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\pi_{\mu_x^{cand}}(X_{iq}^{\mathrm{t}})}{\pi_{\mu_x^t}(X_{iq}^{\mathrm{t}})} &= \prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\frac{1}{X_{iq}^{\mathrm{t}}\sigma_X\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\mathrm{t}}\right)-\mu_X^{cand}\right)^2}{2\sigma_X^2}\right)}{\frac{1}{X_{iq}^{\mathrm{t}}\sigma_X\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\mathrm{t}}\right)-\mu_X^{t}\right)^2}{2\sigma_X^2}\right)} \\ &= \exp\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\frac{(\mu_x^t)^2-(\mu_x^{\mathrm{cand}})^2+2\ln\left(X_{iq}^{\mathrm{t}}\right)\cdot(\mu_X^{\mathrm{cand}}-\mu_X^{\mathrm{t}})}{2\sigma_X^2}\right)\end{aligned}$$

and

$$\begin{aligned}\prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\pi_{\sigma_x^{cand}}(X_{iq}^{\mathrm{t}})}{\pi_{\sigma_x^t}(X_{iq}^{\mathrm{t}})} &= \prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\frac{1}{X_{iq}^{\mathrm{t}}\sigma_X^{cand}\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\mathrm{t}}\right)-\mu_X\right)^2}{2(\sigma_X^{cand})^2}\right)}{\frac{1}{X_{iq}^{\mathrm{t}}\sigma_X^{t}\sqrt{2\pi}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\mathrm{t}}\right)-\mu_X\right)^2}{2(\sigma_X^{t})^2}\right)} \\ &= \prod_{i=1}^{n}\prod_{q=1}^{Q_i}\frac{\sigma_X^t\exp\left(-\frac{\left(\ln\left(X_{iq}^{\mathrm{t}}\right)-\mu_X\right)^2}{2(\sigma_X^{\mathrm{cand}})^2}\right)}{\sigma_X^{\mathrm{cand}}\exp\left(-\frac{\left(\ln\left(X_{iq}^{\mathrm{t}}\right)-\mu_X\right)^2}{2(\sigma_X^{t})^2}\right)} \\ &= \frac{\sigma_X^t}{\sigma_X^{\mathrm{cand}}}\exp\left(\sum_{i=1}^{n}\sum_{q=1}^{Q_i}\frac{\left((\sigma_X^{\mathrm{cand}})^2-(\sigma_X^{\mathrm{t}})^2\right)\left(\ln\left(X_{iq}^{\mathrm{t}}\right)-\mu_X\right)^2}{2(\sigma_X^{\mathrm{cand}})^2(\sigma_X^{\mathrm{t}})^2}\right),\end{aligned}$$

respectively.

Appendix B

Performance of the implemented Bayesian hierarchical approach when accounting for unshared measurement error

In order to test the performance of the implemented Bayesian hierarchical approach when accounting for unshared measurement error, we conducted a simulation study. The results of this simulation study, which are presented in the following, can be found in the appendix of the paper Hoffmann et al. (2017, Radiation Research) "Accounting for Berkson and classical measurement error in radon exposure using a Bayesian structural approach in the analysis of lung cancer mortality in the French cohort of uranium miner".

Results on simulated data sets to illustrate the performance of the proposed approach

We performed a simulation study to illustrate the performance of the Bayesian hierarchical approach to account for exposure uncertainty.

We generated data with an excess relative risk of 1 per 100 working level months and considered lognormal, heteroscedastic and multiplicative measurement error with two exposure periods. For the first exposure period we generated measurement error U with Log(U) following a normal distribution with mean -0.32 and standard deviation 0.8. For the second period the mean was -0.125 and the standard deviation 0.5. For every miner, we generated one exposure value in the first period and one exposure value in the second exposure period. In the first exposure period, cumulated exposure was set equal to the first value and in the second period it was set equal to the sum of the two values.

We used a method based on piecewise exponential variables to generate survival times depending on continuous and time-varying covariates proposed by Hendry (2014).

To test the performance of the Bayesian hierarchical approach to account for Berkson and classical measurement error, 100 data sets were generated for each type of error. Both corrected and uncorrected Excess Hazard Ratio (EHR) estimates were obtained by Bayesian inference under flat prior distributions.

Table S2 Corrected and uncorrected Excess Hazard Ratio (EHR) for classical measurement error and Berkson error under heteroscedastic measurement error. The EHR value to generate the data was 1. Cover probabilities are given for 95% credible intervals.

Type of error	Uncorrected		Corrected	
	EHR per 100 WLM	Cover Probability	EHR per 100 WLM	Cover Probability
Classical error	0.42 [0.17;0.76]	0.20	0.96 [0.39;1.92]	0.95
Berkson error	0.83 [0.42;1.40]	0.90	1.07 [0.50;1.97]	0.95

As can be seen in Table S2, the Bayesian hierarchical approach showed good performance in correcting both Berkson and classical measurement error with estimated cover probabilities of 95% and mean Excess Hazard Ratios close to 1.

Appendix C

Sensitivity of risk estimates on the specification of the piecewise-linear model on the baseline hazard

In order to test the sensitivity of risk estimates on the specification of the piecewise-linear model on the baseline hazard as specified in 5.1.1, we compared the risk estimates obtained on the full cohort and on the post-55 cohort when conducting inference based on the full likelihood and the partial likelihood. The following sensitivity analysis can be found in the appendix of the paper Hoffmann et al. (2017, Radiation Research) "Accounting for Berkson and classical measurement error in radon exposure using a Bayesian structural approach in the analysis of lung cancer mortality in the French cohort of uranium miner".

Comparison of results of the Bayesian hierarchical approach, Partial Likelihood estimation and Full Likelihood estimation

In order to assess the impact of having to specify a form of the baseline hazard, we compared the results for the Bayesian approach under flat prior distributions with the results obtained by using Partial Likelihood (PL) and Full Likelihood (FL) Estimation for the model

$$h_i(t) = h_0(t)\exp\left(\beta_1 X_i^{cum}(t)\right),$$

i.e., the classical form of the Cox proportional hazards model. PL and FL estimation where implemented in R using the survival and the eha package, respectively.

For Bayesian and FL inference, we assumed the piecewise constant model for the baseline hazard

$$h_0(t) = \lambda_j \; \forall t \in I_j = (s_{j-1}, s_j]$$

where $\boldsymbol{\lambda} = (\boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \boldsymbol{\lambda_3}, \boldsymbol{\lambda_4})$. The cut-points of the time-axis $\boldsymbol{s_1}, \boldsymbol{s_2}, \boldsymbol{s_3}, \boldsymbol{s_4}$ were 40, 55, 70 and 85 years. As can be seen in Table 2, the three approaches produce very similar results when it comes to risk estimates and estimated hazard ratios.

Table S1 Comparison of esimates obtained by Bayesian inference, Partial Likelihood Estimation and Full Likelihood Estimation. CI denotes 95% credible intervals for the Bayesian approach and 95% confidence intervals for Partial Likelihood and Full Likelihood Estimation

	β *in* $\mathbf{10^{-2}}$	Hazard Ratio
Bayesian approach		
Full cohort	0.25 [0.12; 0.35]	1.28 [1.14;1.42]
Post-55 cohort	1.18 [0.62; 1.72]	3.27 [1.85; 5.60]
Partial Likelihood Estimation		
Full cohort	0.25 [0.14; 0.36]	1.28 [1.15; 1.44]
Post-55 cohort	1.19 [0.64; 1.74]	3.29 [1.89; 5.73]
Full Likelihood Estimation		
Full cohort	0.24 [0.13; 0.35]	1.30 [1.14; 1.42]
Post-55 cohort	1.27 [0.71; 1.83]	3.57 [2.04; 6.21]

Appendix D

Detailed results on the prior elicitation for average breathing rate for all working conditions

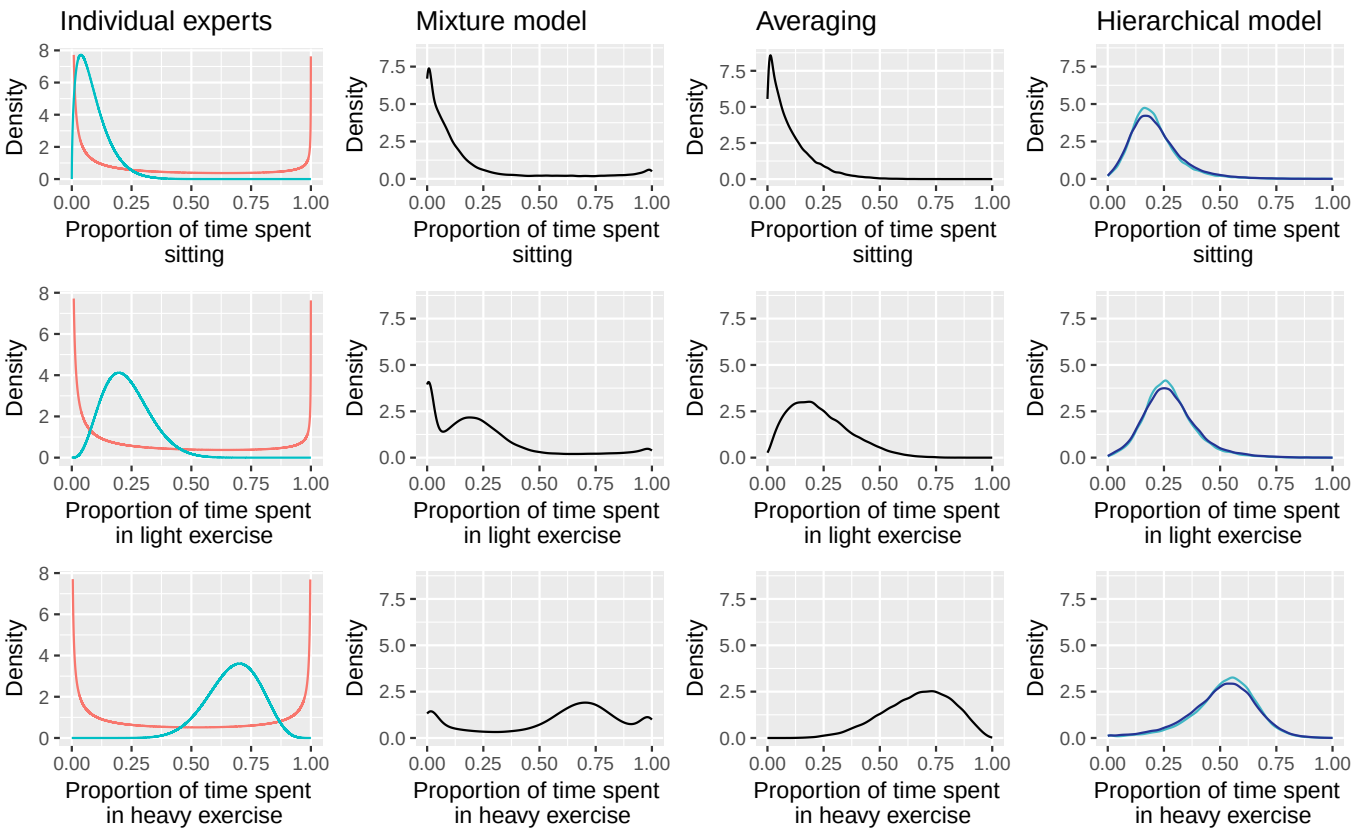

Figure D.1: Hewer before the mechanisation

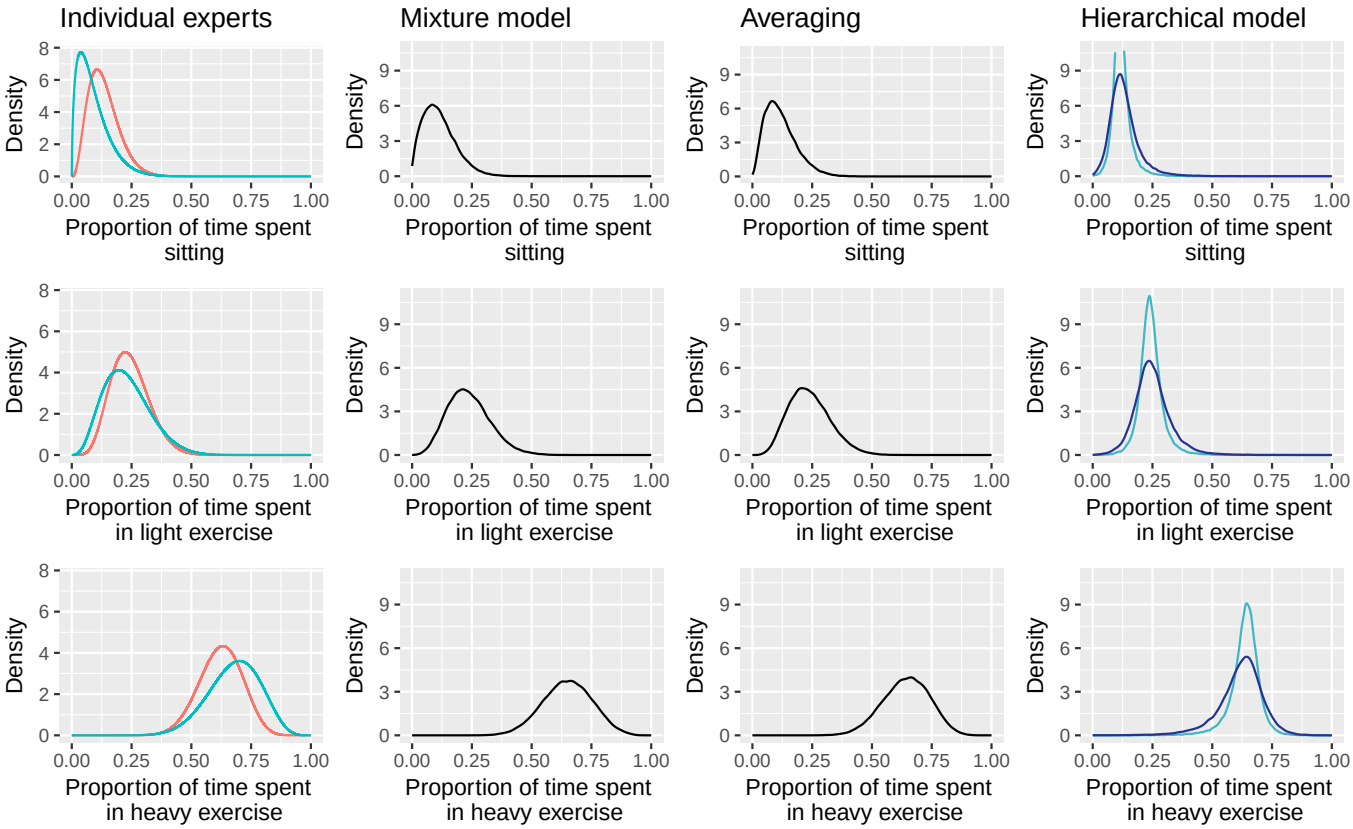

Figure D.2: Underground miner before the mechanisation

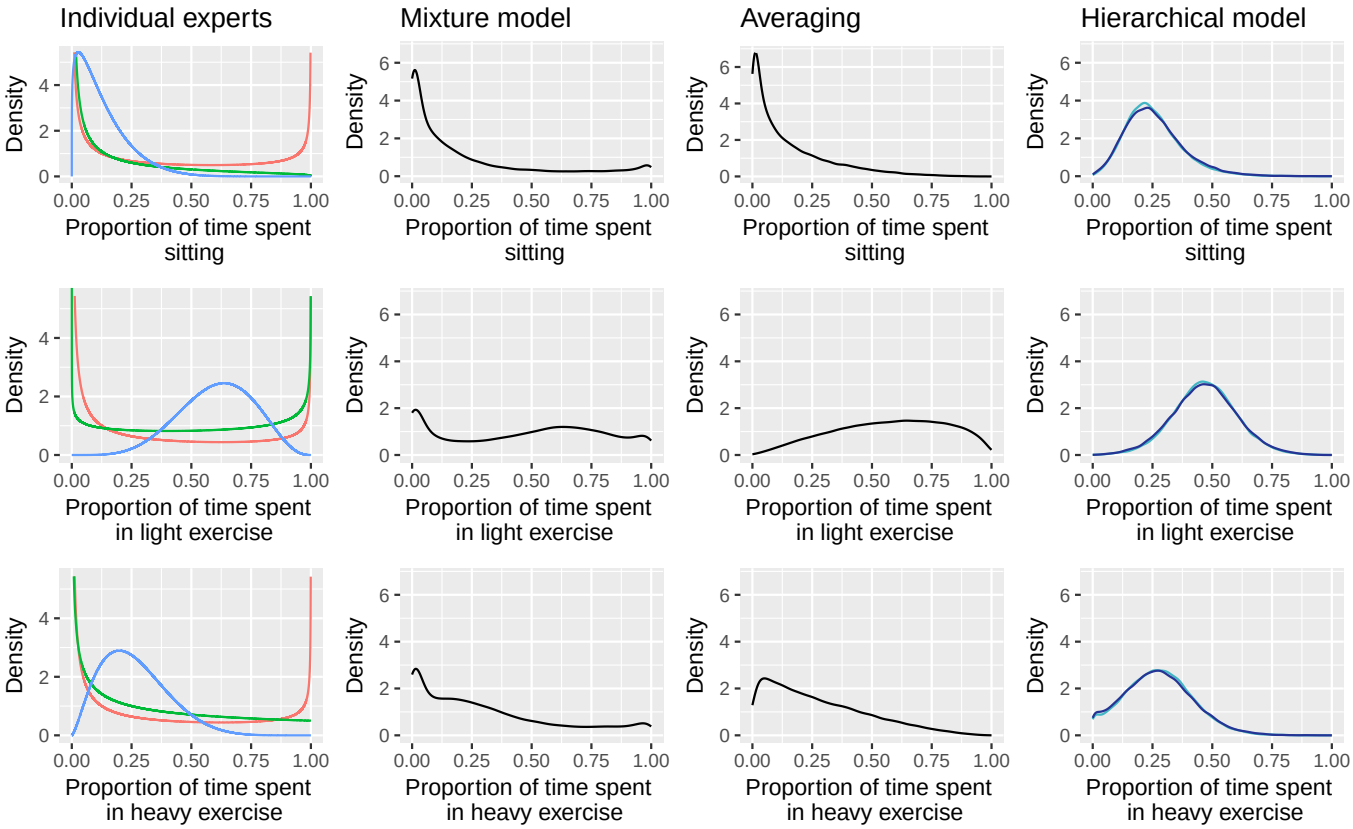

Figure D.3: Underground miner after the mechanisation

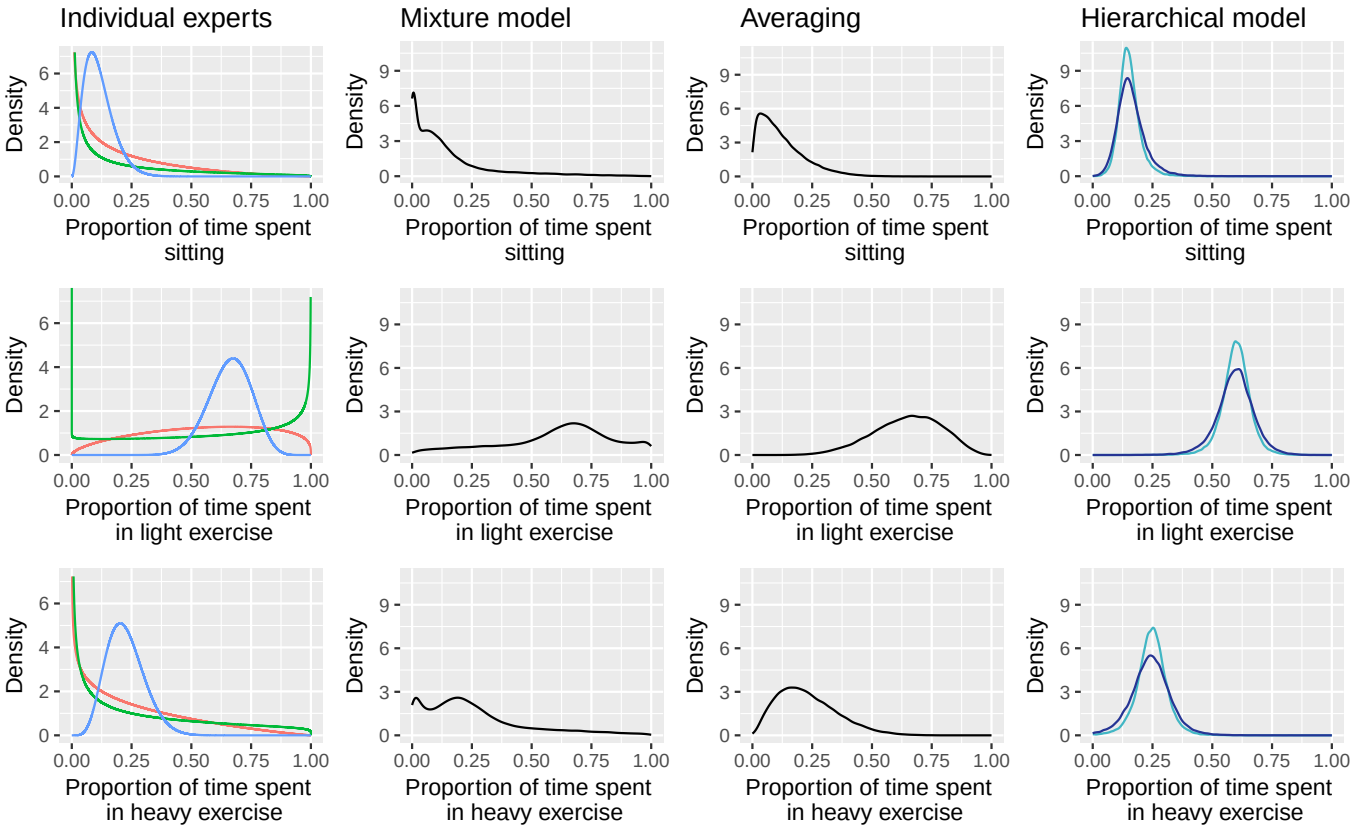

Figure D.4: Open pit after the mechanisation

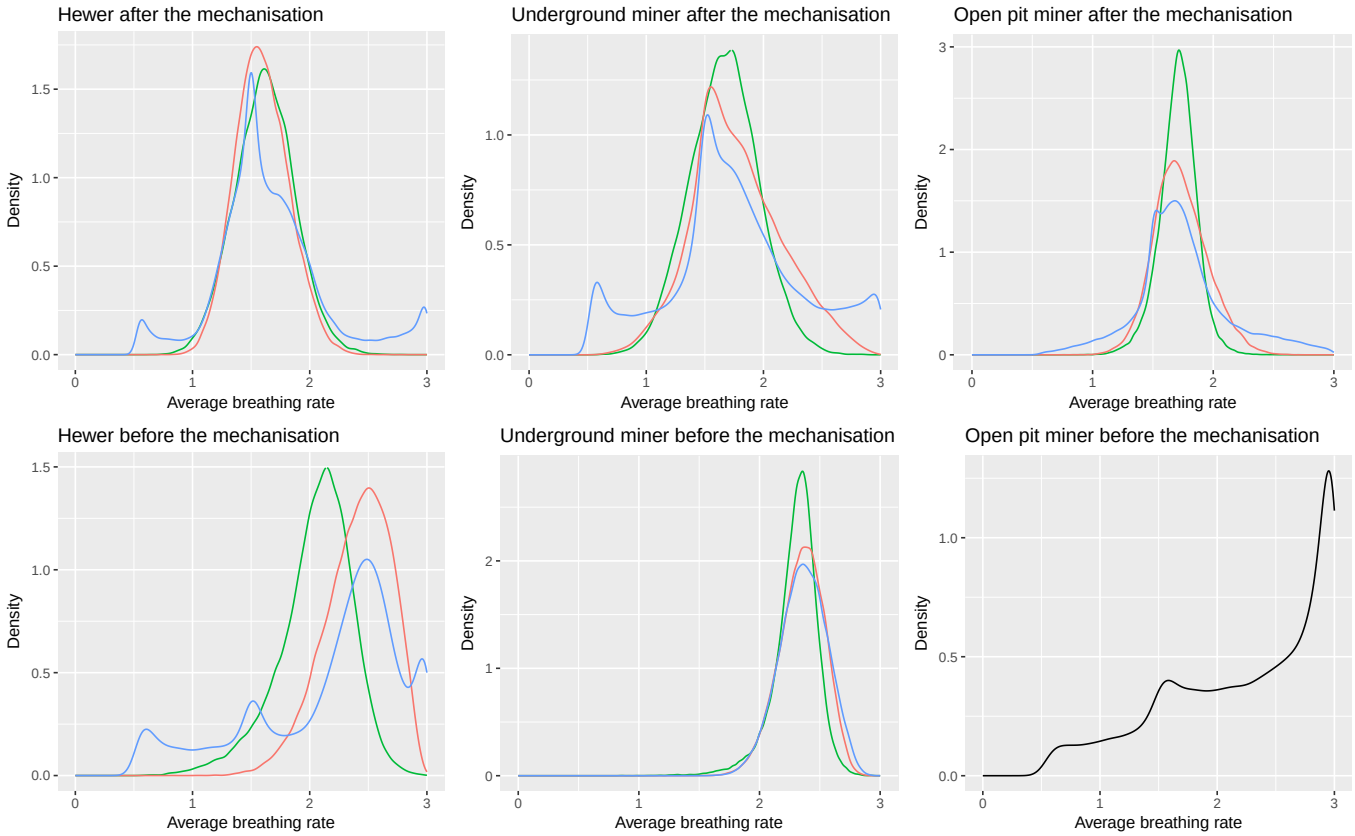

Figure D.5: Summary including a prior on breathing rate for an open pit miner before the mechanisation

Appendix E

Résumé détaillé de la thèse

En épidémiologie des rayonnements ionisants, les erreurs de mesure d'exposition et l'incertitude sur le calcul de la dose absorbée à l'organe constituent des sources d'incertitude importantes entrant en jeu dans la modélisation et l'estimation des risques sanitaires radio-induits et, plus généralement, des relations dose-réponse d'intérêt.

Lorsque les erreurs de mesure d'exposition ne sont pas ou mal prises en compte, elles peuvent mener à des estimateurs de risque biaisés, à une perte de puissance statistique ainsi qu'à une déformation de ces relations dose-réponse. Malgré leurs conséquences délétères et leur omniprésence dans les études épidémiologiques, les erreurs de mesure d'exposition ne sont que très rarement prises en compte dans l'estimation de risques sanitaires. L'une des raisons principales est que les méthodes de correction standards comme la régression calibration ou l'approche par simulation-extrapolation manquent souvent de flexibilité lorsqu'il s'agit de tenir compte de structures complexes d'erreurs de mesure.

Dans les études épidémiologiques menées sur des cohortes professionnelles, par exemple, on s'intéresse souvent à l'association entre l'âge au diagnostic ou l'âge au décès pour une certaine pathologie et l'exposition cumulée à un agent chimique ou physique. Cette exposition cumulée est naturellement dépendante du temps. Dans ce contexte, la prise en compte explicite d'erreurs de mesure d'exposition nécessite de tenir compte de l'historique d'exposition des travailleurs et, en particulier, du fait que les techniques d'évaluation de l'exposition peuvent avoir changé au cours du temps. Ainsi, les expositions reçues dans les premières années de suivi d'une cohorte de travailleurs sont souvent reconstruites rétrospectivement par dires d'experts alors que les expositions reçues dans les années d'exposition plus récentes sont souvent mesurées à l'aide de techniques prospectives et possiblement individuelles. Cela peut conduire à des structures d'erreurs de mesure complexes, caractérisées par des types et magnitudes d'erreurs qui peuvent varier dans le temps. Ainsi, l'utilisation de techniques de mesure prospectives et individuelles de l'exposition peut mener à des erreurs de mesure de type classique. Celles-ci sont par exemple associées au niveau de précision d'un appareil de mesure. Dans le cadre d'une reconstruction rétrospective de mesures d'exposition, il est techniquement impossible de reconstruire l'historique d'exposition de chaque travailleur. Ainsi, il est courant d'estimer puis d'affecter une même valeur d'exposition moyenne à tous les travailleurs associés à un même environnement de travail pour une période calendaire donnée. On parlera d'erreur de type Berkson. L'existence probable d'une erreur d'estimation sur cette exposition moyenne va donc

influencer de la même façon les valeurs d'exposition estimées pour chacun des travailleurs de ce groupe. En outre, les habitudes quotidiennes de chaque travailleur pendant sa journée d'activité ou encore les spécificités des postes qu'un travailleur a occupé pendant son suivi peuvent créer des erreurs de mesure qui se ressemblent sur plusieurs années de suivi consécutives. Ainsi, une reconstruction rétrospective peut donner lieu à des erreurs de type Berkson partagées par plusieurs individus et à des erreurs de type Berkson partagées sur plusieurs années de suivi d'un même individu. Bien que l'impact des erreurs de mesure non-partagées soit désormais bien établi en épidémiologie, celui des erreurs partagées reste très mal connu.

En épidémiologie des rayonnements ionisants, l'estimation d'un risque sanitaire radio-induit peut être non seulement erronée par l'existence d'erreurs de mesure d'exposition mais aussi, dans le cas d'une exposition aux rayonnements ionisants par inhalation ou ingestion, par le fait que ce risque sanitaire puisse être davantage associé à la dose absorbée à l'organe qu'à l'exposition reçue. En effet, la dose absorbée à l'organe dépend non seulement de l'exposition reçue mais aussi, plus généralement, des conditions environnementales associées à cette exposition professionnelle. Comme les paramètres d'entrée des modèles utilisés pour calculer une dose absorbée à l'organe à partir d'une exposition reçue sont rarement mesurables, le calcul de la dose absorbée est lui-même incertain. La prise en compte explicite de cette incertitude dans les estimations de risque sanitaires radio-induits est un défi méthodologique d'actualité.

L'objectif principal de ce travail de thèse est de promouvoir l'utilisation de l'approche hiérarchique bayésienne pour la prise en compte explicite et simultanée des erreurs de mesure d'exposition et de l'incertitude sur le calcul de la dose absorbée à l'organe dans les estimations de risques sanitaires radio-induits dans les études de cohortes professionnelles. En effet, cette approche est reconnue pour sa souplesse et sa pertinence pour la prise en compte de sources d'incertitude multiples et complexes.

Le cas d'étude considéré concerne l'analyse de l'association entre une exposition chronique et à faibles doses au radon (et ses descendants à vie courte) et le risque de décès par cancer du poumon dans la cohorte française des mineurs d'uranium. Le radon, gaz radioactif d'origine naturelle issu de la désintégration de l'uranium 238 contenu dans la croûte terrestre, constitue la principale source d'exposition naturelle aux rayonnements ionisants de l'Homme. Sachant que le radon fut reconnu comme cancérigène pulmonaire en 1988 par le Centre international de Recherche sur le Cancer (CIRC) et que sa concentration devient plus elevée dans des endroits confinés comme les habitations ou les mines, il est important de connaître précisément les risques sanitaires associés afin de permettre in fine des propositions d'amélioration des normes de radioprotection. Dans ce contexte, les cohortes professionnelles de mineurs de fond qui sont, dans le cadre de leur activité professionnelle, eux-mêmes exposés au radon constituent des populations d'étude privilégiées mais pour lesquelles les erreurs de mesure d'exposition au radon ont souvent une structure complexe. Finalement, l'objectif est d'affiner l'estimation actuelle du risque de décès par cancer du poumon associé à une exposition chronique et à faibles doses au radon à partir des données de la cohorte française des mineurs d'uranium en tenant compte de la structure complexe des erreurs de mesure dans cette cohorte et de l'incertitude sur le calcul de la dose absorbée au poumon.

Afin de comparer l'impact de l'existence d'erreurs de mesure d'exposition partagées et non-partagées sur l'estimation d'un risque sanitaire et sur la forme d'une relation exposition-risque,

deux études par simulations ont été conduites dans le contexte d'études de cohortes professionnelles et de l'utilisation de modèles à hasards proportionnels. Différentes structures complexes mais réalistes d'erreurs de mesure (pouvant varier dans le temps) ont été considérées ainsi que différentes valeurs de coefficient de risque et différentes valeurs de variance d'erreur de mesure. Les données d'exposition de la cohorte française des mineurs d'uranium ont été considérées comme valeurs d'exposition vraies ou observées selon le type d'erreur considéré dans les simulations. Les résultats obtenus montrent qu'une incertitude d'exposition partagée sur plusieurs années de suivi d'un même individu conduit à des biais plus élevés dans les estimations de risque ainsi qu'à une déformation plus sévère de la relation exposition-risque qu'une incertitude d'exposition non-partagée ou partagée par plusieurs individus. Ces résultats soulignent l'importance de proposer une caractérisation et une modélisation précise des erreurs de mesure d'exposition potentiellement présentes dans la cohorte française des mineurs d'uranium et d'en tenir compte dans l'estimation du risque de décès par cancer du poumon.

Dans un premier temps, nous avons supposé l'existence d'erreurs de mesure d'exposition exclusivement non-partagées dans la cohorte française des mineurs d'uranium. Un modèle hiérarchique bayésien a été proposé afin de décrire, dans un cadre unique et cohérent, le lien entre l'âge au décès par cancer du poumon d'un mineur et sa vraie exposition au radon (modèle de maladie), puis le lien entre cette exposition vraie et l'exposition observée associée (modèle de mesure avec erreurs Berkson et classique non-partagées) et enfin, dans le cas d'erreurs non-partagées de type classique, l'incertitude sur l'exposition vraie (modèle d'exposition). Un algorithme Monte Carlo par chaînes de Markov de type Metropolis-Within-Gibbs adaptatif a été implémenté en langage Python 2.7 afin de mener l'inférence bayésienne du modèle hierarchique proposé, basé sur une combinaison de sous-modèles probabilistes conditionnellement indépendants (i.e., modèle de maladie, modèle de mesure et modèle d'exposition) et le choix de lois a priori spécifiques sur les différents paramètres inconnus. Une telle approche globale autorise l'estimation conjointe de l'ensemble des paramètres des sous-modèles améliorant ainsi l'estimation de la variance de ces estimations.

Dans un deuxième temps, le modèle hiérarchique bayésien a été complexifié afin de prendre en compte l'existence d'erreurs de mesure d'exposition partagées inter et/ou intra travailleurs dans la cohorte française des mineurs d'uranium puis, dans un troisième temps, l'incertitude sur les paramètres d'entrée d'un modèle simplifié de calcul de la dose absorbée au poumon, élaboré en collaboration avec des dosimétristes. Le calcul de la dose absorbée au poumon fait intervenir des paramètres difficiles à estimer de par l'existence de très peu d'information dans les données. Il est alors crucial d'utiliser d'autres sources d'information à travers la spécification de lois a priori. Dans ce contexte, une élicitation de lois a priori reflétant l'incertitude relative au débit respiratoire moyen d'un mineur d'uranium français dans 6 conditions d'exposition différentes a été conduite auprès de trois experts des conditions d'exposition dans les mines d'uranium françaises. En effet, ce paramètre est notamment un des paramètres d'entrée les plus importants du modèle de calcul de la dose absorbée au poumon.

L'analyse des données de la cohorte française des mineurs d'uranium a montré une augmentation du risque de mortalité par cancer du poumon associé à l'exposition au radon avec la prise en compte d'erreurs de mesure Berkson partagées alors que la prise en compte d'erreurs non-partagées n'a changé l'estimation de risque que marginalement par rapport à une non-prise

en compte de ces erreurs.

En perspective de ce travail, il serait intéressant d'étudier l'impact d'une mauvaise spécification de modèles, que ce soit au niveau de la modélisation de la vraie exposition, de la modélisation des erreurs de mesures ou du modèle de maladie. Enfin, la complexité de l'approche hiérarchique bayésienne nécessiterait d'améliorer leur accessibilité par la mise en place par exemple de logiciels afin de permettre à la communauté scientifique de s'en approprier.

Appendix F

Articles issus de la thèse

RADIATION RESEARCH **187**, 196–209 (2017)
0033-7587/17 $15.00

DOI: 10.1667/RR14467.1

Accounting for Berkson and Classical Measurement Error in Radon Exposure Using a Bayesian Structural Approach in the Analysis of Lung Cancer Mortality in the French Cohort of Uranium Miners

Sabine Hoffmann,[a,1] Estelle Rage,[a] Dominique Laurier,[a] Pierre Laroche,[b] Chantal Guihenneuc[c,2] and Sophie Ancelet[a,2]

[a] *PRP-HOM/SRBE/Lepid, Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France;* [b] *Areva, Direction Santé - 92084 Paris La Défense Cedex, France; and* [c] *EA 4064, Faculté de Pharmacie de Paris, Université Paris Descartes, Sorbonne Paris Cité, Paris, France*

Hoffmann, S., Rage, E., Laurier, D., Laroche, P., Guihenneuc, C. and Ancelet, S. Accounting for Berkson and Classical Measurement Error in Radon Exposure Using a Bayesian Structural Approach in the Analysis of Lung Cancer Mortality in the French Cohort of Uranium Miners. *Radiat. Res.* 187, 196–209 (2017).

Many occupational cohort studies on underground miners have demonstrated that radon exposure is associated with an increased risk of lung cancer mortality. However, despite the deleterious consequences of exposure measurement error on statistical inference, these analyses traditionally do not account for exposure uncertainty. This might be due to the challenging nature of measurement error resulting from imperfect surrogate measures of radon exposure. Indeed, we are typically faced with exposure uncertainty in a time-varying exposure variable where both the type and the magnitude of error may depend on period of exposure. To address the challenge of accounting for multiplicative and heteroscedastic measurement error that may be of Berkson or classical nature, depending on the year of exposure, we opted for a Bayesian structural approach, which is arguably the most flexible method to account for uncertainty in exposure assessment. We assessed the association between occupational radon exposure and lung cancer mortality in the French cohort of uranium miners and found the impact of uncorrelated multiplicative measurement error to be of marginal importance. However, our findings indicate that the retrospective nature of exposure assessment that occurred in the earliest years of mining of this cohort as well as many other cohorts of underground miners might lead to an attenuation of the exposure-risk relationship. More research is needed to address further uncertainties in the calculation of lung dose, since this step will likely introduce important sources of shared uncertainty.

INTRODUCTION

Much of the evidence on the health effects associated with exposure to radon and its decay products (denoted as radon hereafter) originates from epidemiological data on cohorts of occupationally exposed underground miners. Studies conducted on these cohorts have consistently shown a positive association between radon exposure and lung cancer mortality. However, the analyzed risk estimates vary substantially among the eleven most important cohorts of radon-exposed miners (*1*). This variation might be due to factors such as duration of follow-up, background rates of lung cancer, exposure conditions, lifestyle factors, differences in the precision of exposure assessment and simple randomness. Differences in the precision of exposure assessment may lead to varying degrees of exposure uncertainty, which can be expressed as measurement error. Potential consequences of exposure measurement error include biased risk estimates, a distortion of the exposure-risk relationship and a loss in statistical power (*2*). The effects of measurement error on inference depend both on the type and the magnitude of error. Since radon exposure in cohorts of underground miners is assessed with varying degrees of precision, ranging from retrospective exposure estimation based on job-exposure matrices to prospective exposure assessment via personal dosimetry, it is likely that risk estimates will suffer in varying degrees from exposure uncertainty. Owing to changes in radiation protection and in the awareness of the risks associated with radon exposure, we may be confronted with varying degrees of measurement error not only among the different cohorts of miners, but also among different exposure periods in a single cohort.

Despite its pervasive and deleterious consequences, which jeopardize statistical inference, exposure measurement error is only rarely accounted for in studies investigating the association between radon exposure and lung cancer mortality. Moreover, it is questionable whether classical methods that are routinely used to account for exposure measurement error, such as regression calibration and simulation extrapolation (*3*), are versatile enough to

Editor's note. The online version of this article (DOI: 10.1667/RR14467.1) contains supplementary information that is available to all authorized users.

[1] Address for correspondence: IRSN, BP 17, 92262 Fontenay-aux-Roses, France; email: sabine.hoffmann@irsn.fr.

[2] These authors contributed equally to the study.

obtain reliable risk estimates in this context. Indeed, in radon exposure studies of underground miner cohorts, investigators are typically faced with a time-varying exposure variable presenting multiplicative measurement error (*4–6*), where both type and magnitude of error possibly depend on the period of exposure. More importantly, in these cohorts of underground miners, there is generally no validation sample available to estimate the expected value of true exposure, given observed exposure or the true size of measurement error. This makes it difficult to use regression calibration or simulation extrapolation. Regression calibration relies on the expected values of true exposure given observed exposure, which should ideally be estimated on an independent sample. Both regression calibration and simulation extrapolation require the exact knowledge of the magnitude of measurement error and are, in their original version, neither suited for heteroscedastic measurement error[3] nor for measurement error in time-varying variables. Extensions to adapt regression calibration to these situations in the context of survival analysis have recently been proposed (*7*), but their applicability is restricted to cases where a validation sample is available.

In this situation, the Bayesian structural approach, which is arguably the most flexible method to account for exposure uncertainty, makes it possible to estimate all unknown quantities, including risk estimates and true radon exposures, jointly in a unique and coherent framework. This joint modeling allows for the simultaneous estimation of all unknown quantities, where exposure uncertainty is reflected in the estimated precision of risk parameters. Sharing the efficiency of other structural likelihood-based methods for measurement error correction (*8, 9*), this approach can further improve estimations by allowing for the integration of existing subject matter knowledge through the elicitation of prior distributions. This is particularly useful when the available data are poorly informative and strong prior knowledge is available (*10*). Subject matter knowledge can be available either as expert opinion or historical data. For example, in the case of radon-exposed underground miners, it is possible to integrate external information concerning historical data on the conditions in mines or data on similar cohorts of underground miners. When no validation sample is available to estimate the true size of measurement error, prior uncertainty on measurement error variance parameters can be included in Bayesian inference, a situation where classical methods for measurement error correction usually encounter identifiability problems (*2*). Finally, it is difficult or even impossible to calculate confidence intervals for risk estimates obtained by classical methods for measurement error correction, whereas credible intervals are obtainable from Markov Chain Monte Carlo (MCMC) sampling for Bayesian inference.

[3] Heteroscedasticity is defined as the phenomenon in which a variable can show varying degrees of variability depending on the value of a second variable.

In this study, we use a Bayesian structural approach to account for the effects of measurement error on risk estimates obtained from a previously published cohort study of French uranium miners (*11*). Thereby, we model the individual failure time by lung cancer death of each miner, which allows us to account for measurement error in radon exposure on its natural level of occurrence rather than in an aggregated fashion on a stratum level, as is frequently done in Poisson regression. To our knowledge, this is the first time that a Bayesian structural approach is applied to account for exposure measurement error when estimating the risk of lung cancer associated with occupational radon exposure.

MATERIALS AND METHODS

Study Population

The French cohort of uranium miners is a prospective cohort that includes 5,086 males employed as uranium miners for at least one year between 1946 and 1990. Vital status and cause of death were determined until December 31, 2007. At the end of follow-up, 211 miners had died of lung cancer.

Owing to important changes in radiation protection and exposure assessment in French uranium mines in 1956, a post-1955 sub-cohort was defined. This sub-cohort consists of all miners in the original cohort who were first hired after December 31, 1955. It includes 3,377 subjects and presents 94 lung cancer deaths. The post-1955 cohort, the total cohort, the sources of data and the methods of data collection have been described in detail previously (*11*).

Exposure Assessment

Information on radon exposure, expressed in working level months (WLM), was assessed individually for each year of employment, but the method of measurement changed over time (*6*).

Between 1946 and 1955, there was no systematic exposure assessment in French uranium mines. Therefore, the annual radon exposure for this period was retrospectively reconstructed by a group of experts based on environmental measurements performed in the mines and information concerning the miners' type of work and location. In the course of this exposure reconstruction, the experts assumed a monthly exposure value of 1, 2, 5 or 10 working levels for each mining location. This exposure in working levels was then multiplied by the number of months a miner had worked in each mining location to obtain an individual exposure value for each year a miner had worked in this early exposure period.

In 1956, measurements of ambient radon gas concentration at work sites were introduced to estimate monthly exposures for each miner. In this period, exposure levels estimated from grab samples were assigned to work areas. Concurrently, the miners' activity was documented in terms of location and time spent at each location. Individual miner exposures were then calculated by multiplying the time spent in a given work area by its estimated exposure and summing over all work areas. Annual individual records kept in paper archives have been computerized for the purpose of the epidemiological study.

Finally, starting in 1983, personal dosimetry was used to continuously record the potential alpha energy of radon decay products received by each miner. Records were maintained in an electronic database.

Model Formulation

To account for measurement error, we use a structural approach based on conditional probabilistic independence models (*12*) in the

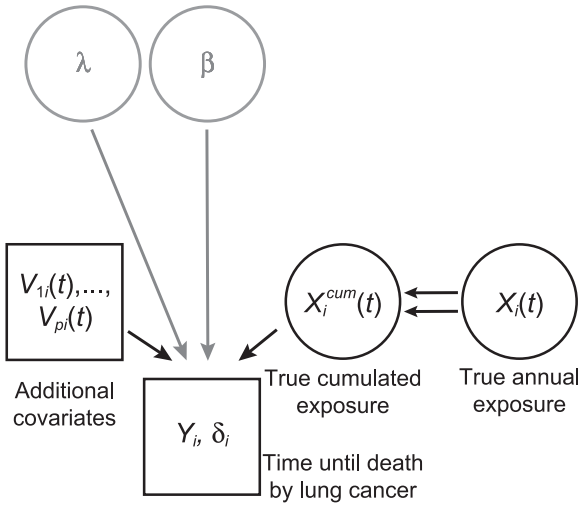

FIG. 1. Directed acyclic graph for the disease model. Circles indicate unknown quantities and rectangles indicate observed variables.

French cohort of uranium miners. In this approach, three submodels must be specified and linked: the disease, measurement and exposure models. The disease model relates the outcome for miner i, $i \in \{1, 2, ..., n\}$ to his true radon exposure, where n is the total number of miners. The measurement model describes the association between observed and true radon exposure, while the exposure model characterizes the distribution of true and unknown radon exposure.

Disease Model

To evaluate the relationship between cumulative occupational radon exposure and lung cancer mortality, we adopted a proportional hazards model, given by

$$h_i(t) = h_0(t)(1 + EHR_i(t)) \quad (1)$$

where $h_i(t)$ is the instantaneous hazard rate of death by lung cancer of miner i at age t, $EHR_i(t)$ is the excess hazard ratio for lung cancer mortality at time t as a function of cumulative radon exposure $X_i^{cum}(t)$ and of potential effect-modifying variables $V_{1i}(t),...,V_{pi}(t)$, which will be specified later. In the absence of other covariates, the baseline hazard $h_0(t)$ corresponds to the hazard of death by lung cancer at time t for an unexposed miner. $X_i^{cum}(t)$ is a time-dependent covariate representing the true and unknown cumulated occupational exposure to radon of miner i at time t, lagged by five years. We assumed a five-year lag to allow for a latency period between radon exposure and lung cancer mortality (*11*).

Age at death by lung cancer of miner i, denoted T_i, is by definition observed only for miners who actually died of lung cancer. Therefore, we define a right censoring variable C_i, which is reserved for miners who died of a cause other than lung cancer, who were still alive on December 31, 2007 or after their 85th birthday, or who were lost to follow-up. We define the outcome of interest by the couple (Y_i, δ_i), where Y_i is time of last news given by $Y_i = min(T_i, C_i)$ and δ_i indicates whether miner i died of lung cancer or not (i.e., $\delta_i = 1$ if $T_i \leq C_i$ and $\delta_i = 0$ if $T_i > C_i$).

An age limitation of 85 years for follow-up was chosen due to the imprecision in determining the exact cause of death in those occurring after the 85th birthday (*13*). Following Heidenreich *et al*. (*5*), who analyzed lung cancer mortality in a large cohort of uranium miners, we assumed that the baseline hazard $h_0(t)$ followed a piecewise constant hazard model. This presents one of the most convenient and popular models for semiparametric survival analysis (*14*). The baseline hazard function can thus be formalized by a finite partition of the time axis $s_0 = 0 < s_1 < s_2 <...< < s_j$ with $s_j = max\ Y_i$ and

$$h_0(t) = \lambda_j \quad \forall t \in I_j = (s_{j-1}, s_j] \quad (2)$$

where I_j denotes the jth time interval ($j = 1,..., J$) and $\lambda = (\lambda_1,... \lambda_j)$ denotes the vector of baseline hazard parameters. In the context of this model, one has to choose cut-points $s_1,..., s_J$ to define the intervals for which the values of the baseline hazard λ_j are assumed to be constant. Heidenreich *et al*. (*5*) chose nine age intervals for this purpose, that is, one for the baseline hazard before 40 years of age, one for the baseline hazard after the age of 75 and intervals of five years in between these two boundaries. We used the same lower boundary of 40 years of age, but decided for a more parsimonious model with steps of 15 years, given the small number of data points in our study compared to Heidenreich *et al*. (*5*). The cut-points of the time axis s_1, s_2, s_3, s_4 we chose were therefore 40, 55, 70 and 85 years and $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$

Potential Effect Modifying Variables

The primary model for $EHR_i(t)$ used here was linear in cumulated exposure with no effect modification, that is $EHR_i(t) = \beta_1 \cdot X_i^{cum}(t)$ (model 1). Vacquier *et al*. (*15*) identified the period of exposure (until 1955 and after 1955) as the most important effect-modifying variable in the French cohort of uranium miners. The corresponding model for the excess hazard ratio is $EHR_i(t) = \beta_1 \cdot X_{i,\leq 1955}^{cum}(t) + \beta_2 \cdot X_{i,>1955}^{cum}(t)$ (model 2), with $X_{i,\leq 1955}^{cum}(t)$ and $X_{i,>1955}^{cum}(t)$ the sum of all radon exposure received by miner i until time t and before 1956 and after 1955, respectively. The associated effect-modifying variable $V_{1i}(l) = \mathbb{I}_{>1955}(X_i(l))$ indicates whether an exposure $X_i(l)$ for miner i at time l was received after 1955 or not. $X_{i,>1955}^{cum}(t)$ can be expressed through $V_{1i}(l)$ by $X_{i,>1955}^{cum}(t) = \sum_{l=l_1}^{t} V_{1i}(l)X_i(l)$, where $X_i(l)$ is the radon exposure a miner i received at time l. Thus, the exposure values for a miner must be summed oveall times $l, l = l_1,...,t$ at which he was exposed to radon before time t. Likewise, $X_{i,\leq 1955}^{cum}(t)$ can be expressed by $X_{i,\leq 1955}^{cum}(t) = \sum_{l=l_1}^{t}(1 - V_{1i}(l))X_i(l)$.

We also chose to include time since exposure as modifying variable. In this context, the risk of radon exposure is estimated by the excess hazard ratio, $EHR_i(t) = \beta_1 X_{i,5-14}^{cum}(t) + \beta_2 X_{i,15-24}^{cum}(t) + \beta_3 X_{i,25+}^{cum}(t)$. (model 3), where $X_{i,5-14}^{cum}(t)$, $X_{i,15-24}^{cum}(t)$ and $X_{i,25+}^{cum}(t)$ are the cumulated radon exposure at time t that miner i received in the last 15 years, 15–25 years or more than 25 years ago, respectively. $X_{i,5-14}^{cum}(t)$ does not include exposures received in the last five years, in consideration of the five-year exposure lag defined earlier. Similar to the period of exposure, two indicator variables, $V_{2i}(l)$ and $V_{3i}(l)$, can be defined to express the effect modification by time since exposure. Note that the time scale of the chosen proportional hazards model is attained age and therefore, the variablesge at exposure and time since exposure contain essentially the same information. Finally, we compared these models with two piecewise linear versions of the linear model with breakpoints at 50 WLM (model 4) or 100 WLM (model 5) to test the linearity of the disease model. The probabilistic dependencies between the variables in the disease model are represented in the directed acyclic graph in Fig. 1.

For the derivation of the likelihood of the disease model see the Supplementary Material (http://dx.doi.org/10.1667/RR14467.1.S1).

Measurement Model

To account for uncertainty in radon exposure assessment, we treated the true radon exposure $X_i(t)$ of miner i at time t as a latent variable (i.e., one that is not observed), while $Z_i(t)$ denotes the observed and error prone measurement of $X_i(t)$. Note that, in general, two different types of measurement error are distinguished in the literature, namely Berkson error and classical measurement error (*2*). In a classical measurement error model, one models the observed variable $Z_i(t)$ conditionally on the latent variable $X_i(t)$, assuming that $Z_i(t)$ is a

function of $X_i(t)$ and of an error term $U_i(t)$, which is independent of $X_i(t)$. Conversely, in a Berkson error model, the true latent variable $X_i(t)$ is modeled conditionally on $Z_i(t)$ and the error term $U_i(t)$ is assumed to be independent of $Z_i(t)$. In general, one assumes a classical error model if an error-prone covariate is measured individually with independent measuring devices and a Berkson error model whenever the value of an error-prone covariate is assigned to a group of subjects. In a Berkson error model, the variability of $X_i(t)$ is assumed to be greater than the variability of $Z_i(t)$ while the opposite is true in a classical error model (*2*).

Consistent with much of the literature on uncertainty in radon exposure (*4–6*), we postulated a lognormal and multiplicative error model to represent the measurement of radon exposure in the French cohort of uranium miners for the complete exposure period. The nature and magnitude of measurement error, however, changed over time.

For the time periods 1946–1955 and 1956–1982, which were characterized by radon exposure estimations by experts and by ambient measurement devices, respectively, we supposed a Berkson error model. This can be justified by the fact that an estimate of radon exposure for a certain position in a mine was assigned to all miners present at a given time and location. The true exposure of a miner might differ from this value, mainly as a result of local and temporal variations in airborne radon concentration, and the resulting error is likely to be independent of observed exposure. We supposed different error variances, expressed through the parameters $\sigma^2_{\varepsilon,1}$,$\sigma^2_{\varepsilon,2}$, $\sigma^2_{\varepsilon,3}$, $\sigma^2_{\varepsilon,4}$ associated with the four distinct time periods identified by Allodji *et al*. (*6*) as characterized by Berkson error, namely 1946–1955, 1956–1974, 1975–1977 and 1978–1982.

We consider the following Berkson measurement error model,

$$X_{iq} = Z_{iq} \cdot U_{iq}, \tag{3}$$

where $log(U_{iq})$ are independent and normal random variables with mean $-\sigma^2_{\varepsilon,p^B_{iq}}/2$ and variance $\sigma^2_{\varepsilon,p^B_{iq}}$, i.e., $\log(U_{iq}) \sim \mathcal{N}(-\sigma^2_{\varepsilon,p^B_{iq}}/2, \sigma^2_{\varepsilon,p^B_{iq}})$ and p^B_{iq} denotes the time period at which Z_{iq} is observed. Note that p^B_{iq} takes values in $\{1, 2, 3, 4\}$ associated with the four different time periods in the Berkson period as defined above.

Using the properties of the lognormal distribution, the variance of measurement error U_{iq} is $V(U_{iq}) = e^{\sigma^2_{\varepsilon,p^B_{iq}}} - 1$. Moreover, its conditional expectation $E(U_{iq}|Z_{iq})$ given the qth observed radon exposure of miner i, Z_{iq} equals 1 $\forall q \in \{1,...,Q_i\}$, where Q_i is the total number of radon exposure measurements for miner i. This implies $E(X_{iq}|Z_{iq}) = Z_{iq}$, and thus a general unbiasedness of the surrogate variable Z_{iq}.

For the time period between 1983 and 2007, on the other hand, exposure measurement was based on personal dosimetry presumably leading to an error best described by a classical error model. Indeed, for this time period, it is reasonable to assume that the true individual radon exposure of each miner was merely assessable with errors that are essentially caused by an inherent variability of the measurement process for each miner i and measurement q. Correspondingly, the classical measurement error model for the fifth time period, 1983–2007, is given by

$$Z_{iq} = X_{iq} \cdot U_{iq}, \tag{4}$$

where $\log(U_{iq}) \sim \mathcal{N}(-\sigma^2_{\varepsilon,5}/2, \sigma^2_{\varepsilon,5})$, reflecting the assumption of homoscedasticity of measurement error in this fifth exposure period, since there were no important changes in the procedure of exposure assessment. As mentioned earlier, the main difference between the Berkson and the classical measurement error model is that U_{iq} is assumed to be independent of Z_{iq} in the former while it is assumed to be independent of X_{iq} in the latter. In the following, we will denote $\sigma_\varepsilon = (\sigma_{\varepsilon,1}, \sigma_{\varepsilon,2}, \sigma_{\varepsilon,3}, \sigma_{\varepsilon,4}, \sigma_{\varepsilon,5})$ the vector of unknown standard deviation parameters of the natural logarithm of measurement error for all periods of Berkson and classical error.

Exposure Model

In the case of Berkson error, the model is fully specified by combining the disease and the measurement model. The correction for classical measurement error, on the other hand, additionally requires the specification of the distribution of true and unknown exposure X_{iq} in the form of an exposure model. As numerous studies on residential and occupational radon found radon exposure to be lognormally distributed (*1, 16, 17*), we assumed a lognormal distribution for X_{iq}. Thanks to changes in radiation protection and to the mechanization of work in the mines, true radon exposure can be assumed to have decreased continuously in the period between 1983 and 2007. To allow different exposure distribution parameters depending on the period of exposure, we modeled mean and standard deviation parameters $\mu_{x,p^c_{iq}}$, and $\sigma_{x,p^c_{iq}}$, of the normal distribution for $log(X_{iq})$ as a function of period of exposure p^c_{iq}. The variable p^c_{iq} takes values in $\{1, 2, 3, 4, 5\}$ corresponding to the exposure periods 1983–1984, 1985–1986, 1987–1989, 1990–1994 and 1995–2007, respectively. These time intervals were chosen so as to obtain a comparable number of measurements in each period.

Therefore, $\mu_x = (\mu_{x,1}, \mu_{x,2}, \mu_{x,3}, \mu_{x,4}, \mu_{x,5})$ and $\sigma^2_x = (\sigma^2_{x,1}, \sigma^2_{x,2}, \sigma^2_{x,3}, \sigma^2_{x,4}, \sigma^2_{x,5})$ denote the vector of unknown mean and variance parameters of the natural logarithm of true exposure.

Combining the Submodels and Defining Prior Distributions

When modeling measurement error using conditional independence models (*13*), one can combine the disease model, the measurement model and the exposure model by using the assumption that measurement error is nondifferential, i.e., disease outcome Y_i and observed exposure Z_{iq} are conditionally independent given true exposure X_{iq} (*2*) for each miner i. The directed acyclic graph for the full hierarchical model, with the results of the combination of the three submodels, is shown in Fig. 2. Independent prior distributions were chosen for the parameters β, λ, μ_x, σ_x and σ_ε.

We adopted centered normal distributions with large variance (10^4) for the regression coefficients $\beta = (\beta_1, \beta_2, \beta_3)$. The prior for β was left truncated to guarantee the positivity of the instantaneous hazard $h_i(t)$ [see Eq. (1)]. Following Ibrahim *et al*. (*14*), we used external data on the national lung cancer mortality rates of the general population for French men between 1968 and 2005 to specify independent and informative gamma priors for λ. In accordance with these mortality rates, we defined gamma priors $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \beta_{0j})$ with α_{0j} taking values 23.66, 35.53, 88.10 and 29.75 and β_{0j} taking values 4.90 10^8, 2.58 10^7, 1.61 10^7 and 3.25 10^6 for the four time intervals, $j = 1, 2, 3, 4$, defined earlier in the model describing baseline hazard λ [see Eq. (2)].

For the parameters $\sigma_{\varepsilon,1}$, $\sigma_{\varepsilon,2}$, $\sigma_{\varepsilon,3}$, $\sigma_{\varepsilon,4}$ and $\sigma_{\varepsilon,5}$ describing the standard deviation of log-transformed measurement error in the five exposure periods, we specified independent normal distributions that were truncated at zero to respect their positivity and centered at the values 0.93, 0.47, 0.42, 0.33 and 0.10, respectively. These values correspond to the estimates that Allodji *et al*. (*6*) obtained for the standard deviation parameters of log-transformed measurement error in the French cohort of uranium miners. The standard deviation parameters of these truncated normal prior distributions were chosen so as to allow for some uncertainty in the values σ_ε, while checking the convergence of the Markov chains to their stationary distribution. Accordingly, these standard deviations were set to 0.03, 0.005, 0.005, 0.005 and 0.0005.

Finally, the prior distributions for the parameters of the exposure model for the classical measurement error period after 1983 were defined in accordance with data on radon exposure in a sub-cohort of the German Uranium Miners cohort consisting of 11,000 miners who were first employed by the Wismut company between 1971 and 1989 (*18*). The miners in this sub-cohort may have been exposed to radon in a way similar to that of the miners in the French cohort after 1983, since their exposure conditions were subject to similar technological developments and the same guidelines for radiation protection.

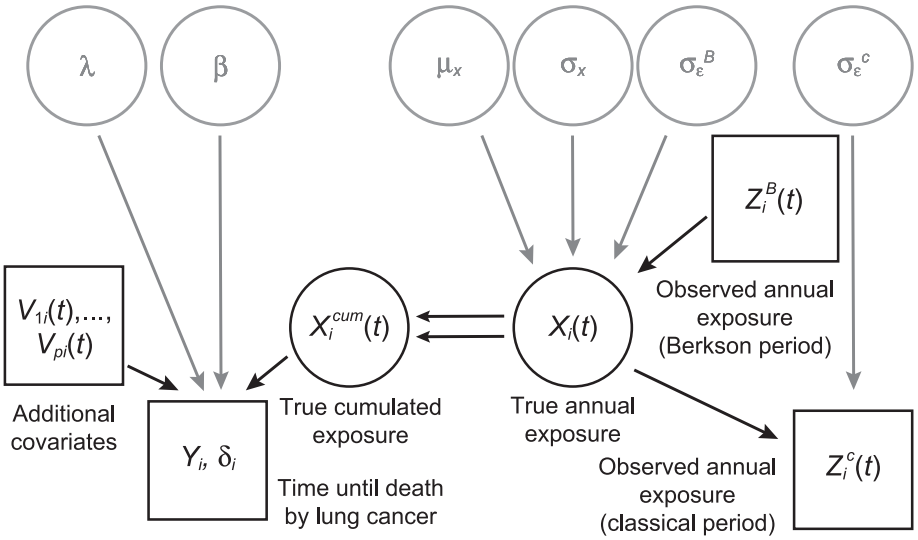

FIG. 2. Directed acyclic graph for the full hierarchical model. Circles indicate unknown quantities and rectangles indicate observed variables. For the sake of clarity, we split the vector σ_ε into $\sigma_\varepsilon^B = (\sigma_{\varepsilon,1}, \sigma_{\varepsilon,2}, \sigma_{\varepsilon,3}, \sigma_{\varepsilon,4})$ for the Berkson error period and $\sigma_\varepsilon^C = \sigma_{\varepsilon,5}$ for the classical measurement error period in this graph.

Among exposed miners in this sub-cohort, the median value of cumulative radon exposure was approximately 5 WLM and mean radon exposure was 9 WLM (*18*). If we divide these values by the average duration of exposure in this cohort, which was seven years, we obtain a yearly median exposure of 0.71 WLM and a yearly mean exposure of 1.29 WLM. Using the properties of the lognormal distribution, we can deduce a mean and a variance of log-transformed exposure values of –0.34 and of 1.18 for this sub-cohort. We therefore assumed the prior expected value of σ_x^2 in the French cohort of uranium miners to be 1.18. With respect to this condition, we chose an inverse gamma prior for σ_x^2 with shape parameter $\alpha_{\sigma_x^2} = 1.75$ and scale parameter $\beta_{\sigma_x^2} = 0.88$ for all $p_c \in \{1, 2, 3, 4, 5\}$. The resulting 95% prior credible interval (95% CI) for σ_x^2 is (0.17; 5.22). Inverse gamma distributions as prior distributions for variance parameters are a classical choice for Bayesian inference to respect their positivity and to obtain analytically tractable full conditional distributions for these parameters.

The steady decrease in annual mean radon exposure, which is observed in cohorts of underground miners after 1965, make the elicitation of the prior distribution for μ_x based on the sub-cohort of the German cohort of uranium miners more difficult. Indeed, all information on radon exposure based on this cohort concerns the period between 1971 and 1989. The exposure model we want to specify, on the other hand, concerns the period between 1983 and 2007 in the French cohort, characterized by classical measurement error. To account for this, we did not directly center the prior distribution for μ_x on its corresponding value observed in the Wismut sub-cohort. This prior is only required to have a non-negligible prior probability for the median radon exposure value to be greater or equal to the median value observed for the Wismut sub-cohort. At the same time, we required that there should be a non-negligible probability for median radon exposure in the French cohort after 1983 to be more than 10 times smaller than in the German cohort to account for the possibility of a gradual decrease in yearly radon exposure in uranium miners. In consideration of these two conditions, we chose $\mu_{x_{p_{iq}^c}} \sim \mathcal{N}(-1.44, 10.24)$ for all p_{iq}^c, corresponding to a 95% prior credible interval of (–7.71; 4.83). The sensitivity of the chosen priors on risk estimates was assessed by comparing the obtained results with results obtained when assuming alternative prior distributions $\mu_{x_{p_{iq}^c}} \sim \mathcal{N}(-1.44, 1.44)$ [95% CI: (–3.79; 0.91)] and $\sigma^2_{x_{p_{iq}^c}} \sim \mathcal{IG}(3.11, 2.47)$ [95% CI: (0.34; 3.70)], where $\mathcal{IG}$ denotes the inverse gamma distribution.

Similarly, we tested the sensitivity of the disease model to the definition of prior distributions for the parameters describing baseline hazard. This was done by running the model with flat uniform distributions between 0 and 200 for the parameters λ_2, λ_3 and λ_4 for both the post-1955 and the total cohort. For the parameter λ_1 describing baseline hazard before 40 years, we always assumed the informative gamma distribution defined earlier, since only one miner died of lung cancer before 40 years in the cohort. Therefore, there was very little information on this parameter in the data and it would have been imprudent to estimate its value based on a flat uniform prior distribution.

Bayesian Inference

Sampling from the joint posterior distribution of the unknown parameters and latent variables was carried out through a MCMC algorithm. More specifically, we adopted an adaptive Metropolis-within-Gibbs algorithm (*19*) to conduct Bayesian inference, as full conditional distributions were analytically intractable with the exception of the full conditional distributions of the parameters of the exposure model. We used component-wise updates, sampling each dimension independently from the others, except for the unknown vector of true exposures X. For this vector, consisting of 47,831 latent exposures, it was unfeasible to update all values at once, insofar as this would lead to a large proportion of samples being rejected and thereby presumably impede convergence. A component-wise updating scheme for the elements of this vector, on the other hand, implies prohibitively expensive computations. Therefore, we used a block-wise sampling scheme for the true exposure X, updating all values for a specific calendar period and a homogenous group of miners in a single step. Groups of miners were determined via a hierarchical ascendant clustering algorithm (*20*), based on values on the principal components obtained by a multiple factor analysis using the variables principal type of mine (underground/open pit), principal location and principal type of job (hewer/other) of each miner. In the adaptive phase, the variances of proposal distributions for all unknown quantities were calibrated so as to obtain acceptance rates of 40% for single components and 20% for vectors to improve the efficiency of the algorithm (*19*). Three chains with different starting values were

TABLE 1
Main Characteristics of the French Cohort of Uranium Miners

Characteristics	Total cohort	Post-1955 cohort
No. of persons	5,086	3,377
Vital status, n (%)		
Alive	2,924 (57.5)	2,412 (71.4)
Deceased (lung cancer)	211 (4.1)	94 (2.8)
Deceased (other cause)	1,724 (33.9)	777 (23.0)
Alive with age $\geq$85 years	187 (3.7)	74 (2.2)
Lost to follow-up	40 (0.8)	20 (0.6)
Age and duration in years, mean (range)		
Age at entry into study	28.8 (16.0–68.0)	28.3 (16.9–57.7)
Age at last observation	64.2 (19.6–85.1)	61.1 (19.6–85.1)
Duration of follow-up	35.4 (0.1–61.0)	32.8 (0.1–51.0)
Radon exposure		
No. of exposed miners, n (%)	4,133 (81.3)	2,910 (86.2)
Cumulated exposure, mean (range)	36.6 (0.003–960.1)	17.8 (0.003–128.4)
Duration of exposure, mean (range)	13.2 (1.0–38.0)	12.9 (1.0–35.0)
No. of miners with cumulative exposure, n (%)		
Exceeding 50 WLM	882 (17.3)	318 (9.4)
Exceeding 100 WLM	378 (7.4)	46 (1.36)
No. of cases with exposure, n (%)		
Exceeding 50 WLM	83 (1.6)	29 (0.8)
Exceeding 100 WLM	42 (0.8)	4 (0.1)

Note. Mean cumulated radon exposure and duration of exposure are calculated only for exposed miners.

run for 125,000 iterations after an initial adaptive phase of 100 cycles for 80 iterations for each of the models. The first 25,000 iterations were discarded as burn-in phase and characteristics of the posterior density were solely based on the remaining samples of the posterior distribution. Trace plots of the chains and the Gelman Rubin statistic (*21*) were used to check convergence for all unknown quantities. Competing disease models were compared via the Deviance Information Criterion (DIC). DIC values are a measure for the fitting abilities of a model with smaller DIC values indicating a better fit to the data (*22*). The adopted algorithm was implemented in Python version 2.7.6.

To assess the performance of our Bayesian hierarchical approach to account for heteroscedastic measurement error of classical or Berkson type, we additionally conducted a simulation study where heteroscedastic measurement error with two exposure periods was generated. To generate failure times depending on time-varying covariates, we used a method proposed by Hendry (*32*). The results of this simulation study are provided in Supplementary Table S2 (http://dx.doi.org/10.1667/RR14467.1.S1).

RESULTS

Description of the French Cohort of Uranium Miners

Table 1 summarizes the main characteristics of the French cohort of uranium miners. Duration of exposure and follow-up were relatively long with a mean duration of exposure of approximately 13 years in both the total and the post-1955 cohort and a mean duration of follow-up of 35.4 years in the former and of 32.8 years in the latter.

Cumulative radon exposure was much higher in the total cohort than in the post-1955 cohort with mean exposures of 36.6 and 17.8 WLM, respectively, and the range of cumulative exposure was much larger in the total cohort than in the post-1955 cohort.

This dramatic decrease in exposure can be explained by important changes in radioprotection (see Fig. 3). Figure 3 shows the annual mean radon exposures for exposed miners is detailed for the three methods of exposure assessment given by the retrospective reconstruction by experts (for the years 1946–1955), ambient measures (for the years 1956–1982) and personal dosimetry (for the years 1983–2007). For a comparison of the distribution of exposure values in the total and in the post-1955 cohort see Supplementary Figs. S1 and S2 (http://dx.doi.org/10.1667/RR14467.1.S1).

Sensitivity to Prior Distributions on Baseline Hazard

Table 2 gives the excess hazard ratio (EHR) and baseline hazard estimates for the total and the post-1955 cohort when assuming the linear model without effect modification and without measurement error correction under different prior distributions for the parameters λ_2, λ_3 and λ_4 describing baseline hazard after 40 years of age.

There was no impact on the EHR point estimate for the total cohort with only slightly narrower credible intervals when assuming informative gamma distributions.

Likewise, the impact on the baseline hazard parameters λ was negligible in the total cohort. Figure 4 shows the impact of the prior choice on the baseline hazard parameters on the posterior distributions of parameters in the post-1955 cohort. When assuming informative prior gamma distributions for baseline hazard, we observe that the data contribute

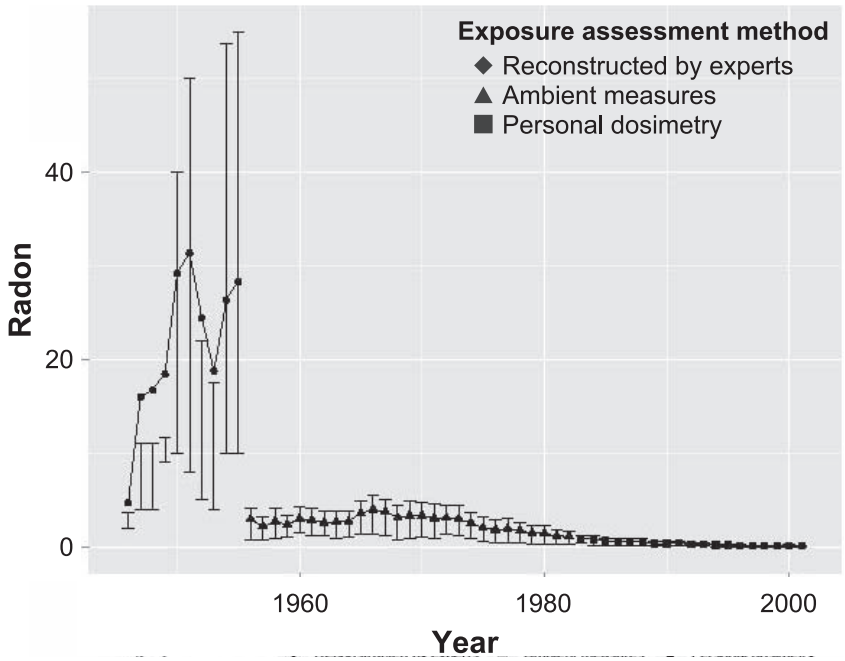

FIG. 3. Annual mean radon exposure of exposed miners (indicated by a diamond, triangle and square depending on the method of exposure assessment) and interquartile range of annual radon exposure (vertical segments) in the French cohort of uranium miners.

to the update of prior beliefs, as the posterior distribution is different from the prior distribution in terms of central location and dispersion. In particular, the ranges of the posterior marginal distributions are smaller, indicating a reduction in uncertainty. In the case of flat uniform prior distributions, the data clearly modify the prior beliefs to obtain much narrower posterior distributions. By comparing the two posterior distributions obtained under the alternative prior choices, we see that the definition of prior distributions on baseline hazard has a notable impact on risk estimates in the post-1955 cohort.

Indeed, we observe a risk estimate of 1.92 per 100 WLM when assuming an informative gamma distribution and a risk estimate of 2.75 per 100 WLM when assuming uniform priors. The 95% credible intervals for risk parameters were much larger in the post-1955 cohort than in the total cohort, reflecting the higher uncertainty on risk parameters due to the smaller sample size and the smaller number of miners who died from lung cancer in this sub-cohort. Under the assumption of uniform priors, all baseline hazard estimates in the post-1955 cohort were smaller than the ones estimated in the full cohort.

Given the sensitivity of the risk estimate in the post-1955 cohort to the choice of informative priors on the baseline hazard parameters and since it is reasonable to believe that the baseline hazard is smaller in an occupational cohort than in the general population, we preferred to base all subsequent analyses on flat uniform baseline hazard priors in the post-1955 cohort. In light of the limited number of miners deceased in this sub-cohort, it seemed reasonable to avoid a too strong influence of the prior distribution on inference. For the total cohort, on the other hand, we kept the hypothesis of informative gamma priors on baseline hazards, as they seemed more appropriate in this cohort.

TABLE 2
Posterior Medians and 95% Credible Intervals for Disease Model Parameters in the Post-1955 Cohort and in the Total Cohort under Model 1 (the Linear Model without Effect Modification) Assuming Different Prior Distributions on Baseline Hazard Parameters

	EHR per 100 WLM	λ_2 10^{-6}	λ_3 10^{-6}	λ_4 10^{-6}
Post-1955 cohort				
Informative gamma priors	1.92 (0.93; 3.22)	1.13 (0.85; 1.48)	5.15 (4.29; 6.13)	8.65 (6.47; 11.29)
Large uniform priors	2.75 (1.18; 5.32)	0.80 (0.45; 1.29)	4.25 (2.85; 6.10)	7.47 (4.60; 11.80)
Total cohort				
Informative gamma priors	0.88 (0.50; 1.36)	1.26 (0.98; 1.58)	5.24 (4.46; 6.11)	9.84 (7.92; 12.13)
Large uniform priors	0.88 (0.44; 1.49)	1.19 (0.82; 1.66)	5.11 (3.96; 6.46)	10.28 (7.72; 13.42)

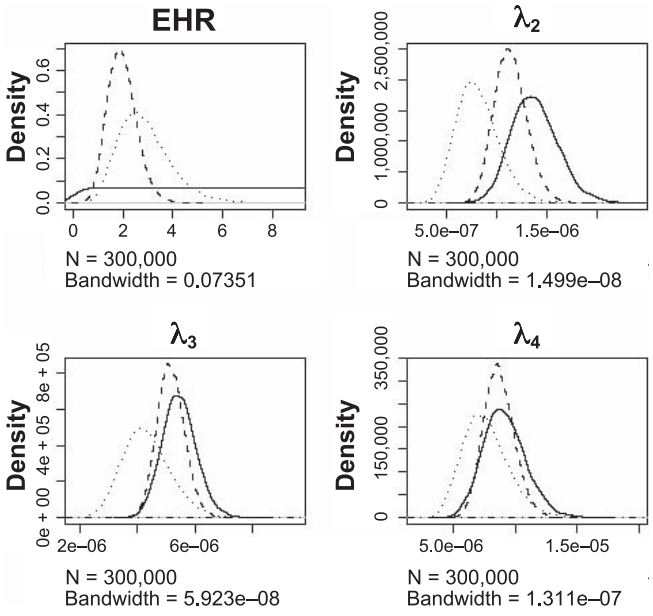

FIG. 4. Posterior distributions for selected parameters of the disease model under flat (dotted line) and informative (dashed line) prior distributions and their corresponding prior distributions (dot-dashed line and solid line, respectively) for the post-1955 cohort.

Indeed, compared to the post-1955 cohort, the use of informative prior distributions on baseline hazard had little or no impact on point estimates, but led merely to slightly more precise estimates.

Accounting for Exposure Uncertainty in the Total Cohort

As described earlier, we implemented different disease models as a function of different effect-modifying variables. Posterior median estimates and 95% credible intervals as well as DIC values for the different disease models for the total cohort are given in Table 3. In general, the comparison of corrected and uncorrected risk estimates in the total cohort reveals no substantial differences in risk estimates or credible intervals.

When assuming a linear excess hazard model without effect modification, the corrected risk coefficient was estimated to be 0.90 per 100 WLM [95% CI: (0.51; 1.41)]. The effect of recent radon exposures (i.e., in the last 15 years) was estimated to induce an excess hazard ratio twice that of exposures received more than 25 years ago. However, DIC values indicate that it was not appropriate to include time since exposure as effect-modifying variable in the total cohort, since this variable did not sufficiently improve the fit of the model to the data. We observe an important attenuation in the exposure-response curve when estimating a piecewise linear model with a risk estimate per 100 WLM that was five times greater for exposures under 50 or 100 WLM than for exposures over these breakpoints. Meanwhile, the risk estimate for exposures received after 1955 was also estimated to be five times greater than the risk estimate for exposures received until 1955. DIC values indicate that period of exposure was the most important effect-modifying variable in the total cohort. These results remain true when accounting for exposure uncertainty. In particular, period of exposure remains the most important

TABLE 3
Posterior Medians, 95% Credible Intervals and Deviance Information Criteria (DIC) for Different Disease Models for the Total Cohort

	Uncorrected		Corrected	
Model	EHR per 100 WLM	DIC	EHR per 100 WLM	DIC
Model 1				
Linear	0.88 (0.50; 1.36)	5,433.37	0.90 (0.51; 1.41)	5,433.30
Model 2				
Period of exposure		5,422.49		5,423.59
until 1955	0.34 (0.04; 0.83)		0.31 (0.02; 0.79)	
after 1955	1.95 (1.16; 2.93)		2.06 (1.34; 3.00)	
Model 3				
Time since exposure		5,435.51		5,435.31
5–15 years	1.49 (0.22; 3.64)		1.88 (0.27; 4.59)	
15–25 years	1.14 (0.12; 2.69)		1.35 (0.17; 3.41)	
≥25 years	0.78 (0.31; 1.40)		0.74 (0.27; 1.57)	
Model 4				
Piecewise linear		5,428.14		5,424.38
<50 WLM	1.99 (0.98; 3.19)		2.12 (1.07; 3.37)	
≥50 WLM	0.37 (0.03; 0.98)		0.34 (0.02; 0.98)	
Model 5				
Piecewise linear		5,426.56		5,422.50
<100 WLM	1.46 (0.84; 2.21)		1.57 (0.92; 2.38)	
≥100 WLM	0.26 (0.01; 0.88)		0.23 (0.01; 0.84)	

TABLE 4
Posterior Medians, 95% Credible Intervals and Deviance Information Criteria (DIC) for Different Disease Models for the Post-1955 Cohort

	Uncorrected		Corrected	
Model	EHR per 100 WLM	DIC	EHR per 100 WLM	DIC
Model 1				
Linear	2.75 (1.18; 5.32)	2,456.09	2.75 (1.20; 5.36)	2,456.09
Model 3				
Time since exposure		2,457.56		2,457.55
5–15 years	2.01 (0.10; 7.91)		2.03 (0.09; 7.86)	
15–25 years	4.82 (1.01; 11.04)		4.82 (1.01; 10.89)	
$\geq$ 25 years	3.20 (0.77; 7.46)		3.19 (0.79; 7.49)	
Model 4				
Piecewise linear		2,457.07		2,456.85
$<$50 WLM	3.20 (0.80; 7.03)		3.22 (0.82; 7.11)	
$\geq$50 WLM	2.82 (0.19; 8.14)		2.72 (0.16; 7.98)	
Model 5				
Piecewise linear		2,457.80		2,458.09
$<$100 WLM	2.90 (1.24; 5.63)		2.85 (1.20; 5.60)	
$\geq$100 WLM	9.28 (0.36; 43.55)		9.50 (0.36; 43.66)	

effect-modifying variable in the full cohort and the attenuation in the exposure-response curve as indicated by the piecewise linear models persists.

The definition of alternative prior distributions for the exposure model did not substantially change corrected risk estimates in the total cohort (results not shown).

Accounting for Exposure Uncertainty in the Post-1955 Cohort

In a subsequent step, we estimated the different disease models with and without measurement error corrections for the post-1955 cohort. Table 4 shows the resulting posterior medians and 95% credible intervals. For the post-1955 cohort, the corrected excess hazard ratio based on the simple linear model was estimated to be 2.75 per 100 WLM [95% CI: (1.20; 5.36)].

In contrast to the full cohort, there was no attenuation in the exposure-response relationship when estimating a piecewise linear model. On the contrary, the risk estimate for exposures over 100 WLM was estimated to be approximately four times greater than for exposures under 100 WLM. This result must be interpreted with great caution, however, since there were only four miners exposed to more than 100 WLM in the post-1955 cohort. Similarly, the uncertainty associated with the risk parameters estimated in the model taking into account time since exposure was very large. Overall, for the post-1955 cohort, DIC values indicate that none of the models including effect-modifying variables describe lung cancer mortality better than the linear model without effect modification. Comparing these results with the estimates obtained when accounting for measurement error, we do not observe substantial differences in risk estimates and in credible intervals in the post-1955 cohort. Similar to the total cohort, the definition of alternative prior distributions for the exposure model did not substantially change corrected risk estimates in the post-1955 cohort.

DISCUSSION

In this analysis of the lung cancer mortality associated with cumulative radon exposure in the French cohort of uranium miners, we built and fitted a Bayesian hierarchical model to explicitly account for exposure uncertainty. We chose a Bayesian hierarchical approach because of its ability to model complex patterns of uncertainty in a coherent and valid inferential framework. This flexibility in modeling stems from the modular nature of the Bayesian hierarchical approach, where submodels are linked by a natural conditional reasoning, which is inherent in the Bayesian conception of probability. In a structure resulting from the conditional linking of submodels, we can arbitrarily complexify one of the submodels without having to modify the other parts of the model and without compromising the validity of inference. Owing to this flexibility, it was possible to account for heteroscedastic and multiplicative measurement error in a time-varying exposure variable, where both type and magnitude of error depend on period of exposure. In contrast to classical functional approaches such as regression calibration, where disjoint steps are used to estimate true exposure and unknown risk parameters, this Bayesian approach allows for joint estimation of true exposure and risk parameters in a unique, global and coherent framework. Consequently, all information available through the observed quantities described by the disease, the measurement and the exposure model, is used to estimate true exposure, while in most classical approaches

only the information available through the measurement model is used for this purpose. In line with this reasoning, other structural approaches that allow for this joint estimation of true exposure and risk parameters have been found to yield more precise estimations than classical functional methods for measurement error correction (*8, 9*). Moreover, structural likelihood-based approaches will generally produce consistent and unbiased risk estimates when accounting for measurement error in proportional hazards models, in contrast to regression calibration, which is neither unbiased nor consistent in this situation (*7*) and which can be even more biased than the naive estimator in complex models (*9*). Compared to other structural likelihood-based approaches, the Bayesian structural approach has the advantage that all uncertainty, including uncertainty in exposure assessment and on error variance parameters, is reflected in credible intervals of risk parameters that are easily obtained by Bayesian inference. Additionally, this approach provides the possibility to include external information on parameters in the form of prior distributions.

Comparison with Other Studies

We modeled measurement error on an individual level using a proportional hazards model and found no substantial impact of Berkson and classical measurement error in the total and in the post-1955 cohort. This is in accordance with results of other occupational cohort studies in which only minimal effect of exposure and dose uncertainty on risk estimates were found (*5, 23, 24*). In a published collaborative analysis of individual data from 13 case-control studies on residential radon, however, Darby *et al.* (*25*) adjusted for uncertainties in radon exposure and observed a notable increase in relative risk estimates for lung cancer. This difference may be explained by two important factors that have a major impact on the influence of exposure uncertainty on risk estimates, namely the measurement error variance and the variability of true exposure. In the case of additive classical measurement error in linear regression with a continuous dependent variable, it is straightforward to show that when there is only one predictor in the model, its estimated coefficient will be attenuated by the factor $\sigma_x^2/(\sigma_\varepsilon^2 + \sigma_x^2)$, where σ_x^2 is the variance of the predictor X and σ_ε^2 is the variance of measurement error (*2*). When modeling binary outcomes or failure times, which present outcomes that are undoubtedly of more interest in radiation epidemiology, there is in general no simple expression for this attenuation factor. However, the magnitude of measurement error and the variability of true exposure seem to act in the same way when it comes to the impact of exposure uncertainty (*16*).

In case-control studies on residential radon, exposure uncertainty will likely be greater than in occupational cohort studies, because exposure assessment is mostly performed retrospectively and few measurements are taken to reconstruct the exposure history of several decades. In occupational cohort studies, on the other hand, there is often a prospective and direct exposure assessment, at least for the more recent exposure periods. At the same time, the variability of radon exposure itself is much smaller in studies on residential radon than in occupational studies on cohorts of underground miners. The exposure variance in occupational cohort studies might be further amplified through the accumulation of annual exposure values. Indeed, under the assumption of a positive correlation between the annual exposure values of a miner, the variance of cumulated exposure is greater than the sum of the variances of annual exposure. Thus, this accumulation will probably make the results on the health effects of radon exposure that are based on cohorts of uranium miners even less vulnerable to the impacts of exposure uncertainty. In light of these facts, it seems important to complement the information obtained by case-control studies on residential radon by extrapolating results on the association between radon exposure and lung cancer mortality obtained on cohorts of occupationally exposed underground miners. In comparison to most other cohorts of underground miners, the French cohort of uranium miners is characterized by low levels and low rates of radon exposure for a long period of time and prospective exposure assessment of good quality, making it especially valuable for the definition of radioprotection standards.

Difference of Retrospective Exposure Assessment and Prospective Exposure Assessment

The only exposure period characterized by retrospective exposure assessment in the French cohort of uranium miners was the period before 1956 for which only few prospective measurements of radon exposure were available. Unfortunately, we cannot exclude that the categorization made in exposure reconstruction for these early years led to differential measurement error (*26*). Moreover, there might be shared uncertainty among miners at the same mining location and an auto-correlative structure of measurement error. There is no conventional method for measurement error correction that enables modeling of this potentially shared, correlated and differential measurement error, since virtually all available methods rely on the assumption of nondifferential measurement error, at least in the case when no validation sample is available (*2*). It would be possible to build a Bayesian hierarchical model reflecting this complex error structure. However, this model would inevitably rely on many hypotheses, for example, on the autocorrelative structure of measurement error for several exposure values of a miner or on the magnitude of nondifferentiality. These hypotheses are very difficult or even impossible to corroborate for an error that occurred more than 60 years ago. In this situation, one could elaborate on different scenarios with varying degrees of differential, shared and correlated exposure uncertainty for this period and calculate corrected risk estimates under these

scenarios. The elaboration of these scenarios was out of the scope of our study.

In the French cohort of uranium miners, we found period of exposure to be the most important effect-modifying variable. The most plausible explanation for this effect modification is the difference in quality of exposure assessment marked by retrospective exposure reconstruction until 1955 and prospective exposure assessment of good quality after 1955. Moreover, the resulting association of high exposure rate and low-quality exposure assessment seems to create an attenuation in the exposure-response relationship for high-cumulative radon exposures, a phenomenon that has been observed in numerous occupational cohort studies (*27*) and previously reported in analyses of the lung cancer mortality of uranium miners (*28*). The fact that this attenuation, as well as the effect modification by period of exposure, persisted when accounting for measurement error in the total cohort might indicate that other mechanisms are responsible for this distortion of the exposure-response relationship. In light of the complex structure of measurement error caused by the retrospective exposure assessment until 1955, however, it is more plausible to assume that it was impossible to completely correct for the effects of measurement error in this early period. Moreover, when restricting our analysis to the post-1955 cohort, which is characterized by prospective exposure measurements of good quality, we do not observe any attenuation in the dose-response curve for high cumulative exposure values. This finding supports the hypothesis that this nonlinearity is caused by exposure uncertainty, rather than by other mechanisms that have been proposed in this context, such as confounding risk factors, a healthy worker survivor effect or the saturation of biological pathways (*27*). Indeed, if any of the latter factors were responsible for the attenuation of the exposure-response relationship in the French cohort of uranium miners, one would expect to observe the same attenuation in the post-1955 cohort as in the total cohort.

Ignoring the distortion of the exposure-risk relationship that might be introduced by the difference in quality of exposure assessment presents a potentially precarious and jeopardous choice. Indeed, if there is a resulting leveling off for high exposures, a linear model will substantially underestimate the risk for low exposures. This further stresses the importance of the prospective exposure assessment of good quality in the period after 1955 in the French cohort of uranium miners. In future studies on this cohort, we recommend separate analyses on the total cohort and the post-1955 cohort when accounting for measurement error. Unfortunately, the modeling of effect modification in the post-1955 cohort is possible only to a limited extent because of the reduced number of miners and therefore, the reduced number of lung cancer deaths observed in this subcohort. This is reflected by the large credible intervals we observed, such as for effect modification by time since exposure. Since the information available in the disease model is involved in the estimation of true exposure, it is nevertheless important to choose a well-fitting disease model from which all deviations might be considered as measurement error. With that in mind, one could attempt to incorporate model uncertainty in the hierarchical Bayesian approach proposed here by considering several disease models simultaneously using a Bayesian model averaging framework (*29*).

The examination of the impacts of retrospective exposure assessment and the resulting differences in exposure uncertainty are crucial when analyzing the lung cancer mortality associated with occupational radon exposure, since all of the eleven most important cohorts of underground miners partially or fully rely on retrospective radon exposure assessment (*1*). Moreover, in most of these cohorts, the quality of exposure assessment improved with time while the level of radon exposure decreased due to the introduction of exposure limits and radiation protection guidelines. The importance of the quality of exposure assessment has been discussed previously in the analysis of lung cancer mortality in the Czech and the French cohort of uranium miners (*30*).

Limitations

While measurement error occurring in the exposure period after 1983 seems to be adequately described by unshared multiplicative classical error, the period between 1956 and 1982, which was characterized by ambient measurements, still presents some challenges. These challenges arise primarily because the measurement error that occurred in this period might neither be completely independent among miners nor purely Berkson. Even though this assumption is likely to hold for the most important error component in this period, namely local variations in radon concentration (*6*), there are other measurement error components for which this hypothesis is far more questionable. The precision of the measurement device used for ambient measurements, for instance, is more likely to create a measurement error component that is shared among miners who worked at the corresponding work location in the week the measurement was taken. If we had information on the exact work locations of miners today, it would be straightforward to introduce different parameters according to these groups of miners, in whom components of exposure uncertainty in the same week of exposure assessment are shared, thus facilitating a coherent hierarchical measurement model. In this framework, the distribution of the error occurring for the qth measurement of miner i in the jth group of miners sharing uncertainty could then be written as $[U_{iqj}] = [U_{iqj}|c_j][c_j]\, dc_j$, where c_j presents the shared error of group j. Because of its flexibility, the Bayesian hierarchical approach would allow us to couple this hierarchical measurement model with the disease model, as described earlier. Despite this possibility, it was not feasible to do so in our analyses, since no

information on the detailed presence of miners at work stations is available today. Alternative approaches to account for shared error components are unlikely to resolve this problem, since they would also require detailed information on the groups of miners who share a common source of error. Importantly, in this context it is common to treat this mixture of shared classical and unshared Berkson error as pure unshared Berkson error for the sake of simplicity. Yet, one may assume that almost all observational studies that assume pure Berkson error are in reality confronted with an additional component of shared error. The occurrence of pure Berkson error is presumably quite rare, since it would imply that the value that is assigned to a group of subjects is precisely known. This is a reasonable assumption in experimental settings when there is a prefixed value, for example a known radiation dose delivered by a machine, and individual dose values deviate from this value due to the imprecision in the realization of the experiment. In the majority of observational studies in epidemiology, however, where a purported mean value is estimated or measured for a group of subjects, it is unrealistic to assume that this assigned value is precisely known. In the French cohort of uranium miners, considering this error component as unshared rather than shared uncertainty can be assumed to have only a negligible impact on risk estimates and their corresponding credible intervals. Indeed, in the period 1956–1982, there were typically more than 30 measurements of radon exposure per person-year. The fact that exposures were first summed over work locations to obtain weekly exposures, which were then again summed to obtain yearly exposures, will likely attenuate the impact of this shared error component. In line with this reasoning, Stayner *et al.* (*24*) and Little *et al.* (*23*) observed only a minor impact of shared uncertainty in cumulative occupational radiation exposure in nuclear workers and in radiologic technologists, respectively.

Another limitation of this study concerns the assumptions, which must be made when using a structural likelihood-based approach to correct for measurement error in survival analysis. In general, the unbiasedness of this type of approach comes at the price of making additional model assumptions on the distribution of true exposures. In survival analysis, the use of a likelihood-based approach also implies that one must specify the full likelihood which entails the specification of the form of the baseline hazard function. In the French cohort of uranium miners, we assumed a piecewise constant baseline hazard, based on data on the general lung cancer mortality in the French male population. We tested the influence of this modeling choice under the Cox proportional hazards model in its classical form $h_i(t) = h_0(t)\exp(\beta_1 X_i^{cum}(t))$, since partial likelihood estimates are easily obtained by standard statistical software for this model. The comparison of risk estimates obtained by the proposed Bayesian approach with risk estimates obtained by maximizing the partial likelihood showed that there were virtually no differences in estimates in the full and in the post-1955 cohort (see Supplementary Table S1; http://dx.doi.org/10.1667/RR14467.1.S1). The relative robustness of risk estimates to this modeling assumption is also reflected by the fact that we found uncorrected risk estimates similar to Rage *et al.* (*11*), who estimated excess relative risk in the French cohort of uranium miners by means of Poisson regression using a large number of strata.

Implications for Radiation Protection

Recommendations for radiological protection against radon are based considerably on the health effects of occupational radon exposure observed in cohorts of uranium miners (*31*). Therefore, it is important to refine the estimation of the lung cancer risk associated with radon exposure in uranium miners by accounting for measurement error. To further elucidate the effects of radon exposure, the International Commission for Radiation Protection recommends the use of a dosimetric approach by calculating radiation doses to the respiratory tract (*31*). The calculation of these radiation doses will imply further uncertainties, which should be accounted for in order to derive credible intervals that properly reflect both the statistical variability of the data and the incomplete knowledge of radiation dose (*1*).

Finally, it seems important to adjust for exposure to gamma rays and long-lived radionuclides in the association between radon exposure and lung cancer, since these variables present a major potential for confounding. Indeed, these sources of ionizing radiation are highly correlated in miners and each substantially contributes to the total lung dose (*13*). Modeling individual radiation doses for all exposures in a common disease model could be possible in the French cohort of uranium miners, since they were assessed via ambient measurement and personal dosimetry beginning in 1956. However, gamma rays and long-lived radionuclides doses are likely to present measurement error and dose uncertainties as well. Entering these variables in the model would complicate the effects of measurement error and dose uncertainty. The risk estimate for radon exposure would be further contaminated by the errors in these confounding variables, and it is impossible to predict whether this will artificially attenuate or inflate the risk estimate for radon exposure. Therefore, one must model dosimetric uncertainties for the three exposures simultaneously to obtain an unbiased estimate for radon exposure. To be able to include doses from all three exposure variables, it is necessary to reduce the correlation in the three variables. One way to achieve this could be to add more precise information on parameters that intervene in the dose calculation for one or two exposure variables, but not for all of them.

Because of its flexible and modular nature, the Bayesian hierarchical modeling approach presents a promising framework to simultaneously account for measurement error and other dose uncertainties for all three radiation

exposures in a unique and coherent framework. The possibility to include additional information on uncertain parameters that intervene in the dosimetric model through the prior elicitation by experts makes this approach even more appealing in this context. The post-1955 French uranium miner cohort is characterized by a long duration of follow-up, good-quality prospective exposure assessment for radon, gamma rays and long-lived radionuclides, as well as low levels of cumulative exposure in general. Thus, this cohort is highly suitable for these future analyses that account for dose uncertainties in all three sources of ionizing radiation.

SUPPLEMENTARY INFORMATION

Fig. S1. Distribution of cumulative exposure values for miners employed until 1955 and after 1955.

Fig. S2. Distribution of cumulative exposure values under 300 WLM for miners employed until 1955 and after 1955.

Table S1. Comparison of estimates obtained by Bayesian inference, partial likelihood estimation and full likelihood estimation.

Table S2. Corrected and uncorrected EHR for classical measurement error and Berkson error under heteroscedastic measurement error. Data was generated using an EHR value of 1. Cover probabilities are given for 95% credible intervals.

ACKNOWLEDGMENTS

This work was partially supported by AREVA-NC in the framework of a bilateral agreement between the IRSN and AREVA-NC. We thank the two anonymous reviewers and the associate editor for their constructive comments.

Received: March 21, 2016; accepted: November 10, 2016; published online: January 24, 2017

REFERENCES

1. Sources-to-effects assessment for radon in homes and workplaces. UNSCEAR 2006 Report to the General Assembly, with scientific annexes, Vol. II, Effects of ionizing radiation. New York: United Nations Scientific Committee on the Effects of Atomic Radiation; 2006. p. 197–334.
2. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. Boca Raton, FL: Chapman Hall; 2006.
3. Cook JR, Stefanski LA. Simulation-extrapolation estimation in parametric measurement error models. J Am Stat Assoc 1994; 89:1314–28.
4. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. Stat Med 2005; 24:1713–23.
5. Heidenreich W, Tomasek L, Grosche B, Leuraud K, Laurier D. Lung cancer mortality in the European uranium miners cohorts analyzed with a biologically based model taking into account radon measurement error. Radiat Environ Biophys 2012; 51:263–75.
6. Allodji RS, Leuraud K, Bernhard S, Henry S, Bénichou J, Laurier D. Assessment of uncertainty associated with measuring exposure to radon and decay products in the French uranium miners cohort. J Radiol Prot 2012; 32:85–100.
7. Liao X, Zucker DM, Li Y, Spiegelman D. Survival analysis with error-prone time-varying covariates: A risk set calibration approach. Biometrics 2011; 67:50–8.
8. Schafer DW, Purdy KG. Likelihood analysis for errors-in-variables regression with replicate measurements. Biometrika 1996; 83:813–24.
9. Kuchenhoff H, Carroll RJ. Segmented regression with errors in predictors: semi-parametric and parametric methods. Stat Med 1997; 16:169–88.
10. Albert I, Donnet S, Guihenneuc-Jouyaux C, Low-Choy S, Mengersen K, Rousseau J, et al. Combining expert opinions in prior elicitation. Bayesian Anal 2012; 7:503–32.
11. Rage E, Caer-Lorho S, Drubay D, Ancelet S, Laroche P, Laurier D. Mortality analysis in the updated French cohort of uranium miners (1946–2007). Int Arch Occup Environ Health 2015; 88:717–30.
12. Richardson S, Gilks WR. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. Am J Epidemiol 1993; 138:430–42.
13. Rage E, Vacquier B, Blanchardon E, Allodji RS, Marsh JW, Caer-Lorho S, et al. Risk of lung cancer mortality in relation to lung doses among French uranium miners: Follow-up 1956–1999. Radiat Res 2012; 177:288–97.
14. Ibrahim JG, Chen MH, Sinha D. Bayesian survival analysis. New York: Springer; 2001.
15. Vacquier B, Rogel A, Leuraud K, Caer S, Acker A, Laurier D. Radon-associated lung cancer risk among French uranium miners: modifying factors of the exposure-risk relationship. Radiat Environ Biophys 2008; 48:1–9.
16. Heid I, Kuchenhoff H, Wellmann J, Gerken M, Kreienbrock L, Wichmann HE. On the potential of measurement error to induce differential bias on odds ratio estimates: an example from radon epidemiology. Stat Med 2002; 21:3261–78.
17. Zettwoog P. State-of-the-art of the alpha individual dosimetry in France. In: Gomez M, editor. International conference radiation hazards in mining: Control, measurements and medical aspects, Oct. 4–9, 1981 Colorado School of Mines, Golden, CO. New York: Society of Mining Engineers of the American Institute of Mining, Metallurgical, and Petroleum Engineers; 1981. p. 4–9.
18. Kreuzer M, Grosche B, Dufey F, Schnelzer M, Tschense A, Walsh L. The German uranium miners cohort study (Wismut cohort), 1946–2003. Bundesamt fur Strahlenschutz; 2011.
19. Roberts GO, Rosenthal JS. Examples of adaptive MCMC. J Comput Graph Stat 2009; 18:349–67.
20. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. Vol. 1, Springer series in statistics. Berlin: Springer; 2001.
21. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci 1992:457–72.
22. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat Methodol 2002; 64:583–639.
23. Little MP, Kwon D, Doi K, Simon SL, Preston DL, Doody MM, et al. Association of chromosome translocation rate with low dose occupational radiation exposures in U.S. radiologic technologists. Radiat Res 2014; 182:1–17.
24. Stayner L, Vrijheid M, Cardis E, Stram DO, Deltour I, Gilbert SJ, et al. A Monte Carlo maximum likelihood method for estimating uncertainty arising from shared errors in exposures in epidemiological studies of nuclear workers. Radiat Res 2007; 168:757–63.
25. Darby S, Hill D, Auvinen A, Barros-Dios J, Baysson H, Bochicchio F, et al. Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 European case-control studies. BMJ. 2005; 330:223.
26. Flegal KM, Keyl PM, Nieto FJ. Differential misclassification

arising from nondifferential errors in exposure measurement. Am J Epidemiol 1991; 134:1233–46.

27. Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environ Health 2003; 29:317–24.

28. Hunter N, Muirhead CR, Tomasek L, Kreuzer M, Laurier D, Leuraud K, et al. Joint analysis of three European nested case-control studies of lung cancer among radon exposed miners: exposure restricted to below 300 WLM. Health Phys 2013; 104:282–92.

29. Hoeting JA, Madigan D, Raftery A, Volinsky, CT. Bayesian model averaging: a tutorial. Stat Sci 1999; 382–401.

30. Tomasek L, Agnes R, Tirmarche M, Mitton N, Laurier D. Lung cancer in French and Czech uranium miners: radon-associated risk at low exposure rates and modifying effects of time since exposure and age at exposure. Radiat Res 2008; 169:125–37.

31. Tirmarche M, Harrison J, Laurier D, Paquet F, Blanchardon E, Marsh JW. Lung cancer risk from radon and progeny and statement on radon. Ann ICRP 2010; 40:1–64.

32. Hendry DJ. Data generation for the Cox proportional hazards model with time-dependent covariates: A method for medical researchers. Stat Med 2014; 33:436–54.

Shared and unshared exposure measurement error in occupational cohort studies and their effects on statistical inference in proportional hazards models

Sabine Hoffmann[1*], Dominique Laurier[1], Estelle Rage[1], Chantal Guihenneuc[2☯], Sophie Ancelet[1☯]

1 Institut de Radioprotection et de Sûreté Nucléaire, Fontenay-aux-Roses, France

2 Faculté de Pharmacie, Université Paris Descartes, Paris, France

☯These authors contributed equally to this work.

* sabine.hoffmann@irsn.fr

Abstract

Exposure measurement error represents one of the most important sources of uncertainty in epidemiology. When exposure uncertainty is not or only poorly accounted for, it can lead to biased risk estimates and a distortion of the shape of the exposure-response relationship. In occupational cohort studies, the time-dependent nature of exposure and changes in the method of exposure assessment may create complex error structures. When a method of group-level exposure assessment is used, individual worker practices and the imprecision of the instrument used to measure the average exposure for a group of workers may give rise to errors that are shared between workers, within workers or both. In contrast to unshared measurement error, the effects of shared errors remain largely unknown. Moreover, exposure uncertainty and magnitude of exposure are typically highest for the earliest years of exposure. We conduct a simulation study based on exposure data of the French cohort of uranium miners to compare the effects of shared and unshared exposure uncertainty on risk estimation and on the shape of the exposure-response curve in proportional hazards models. Our results indicate that uncertainty components shared within workers cause

more bias in risk estimation and a more severe attenuation of the exposure-response relationship than unshared exposure uncertainty or exposure uncertainty shared between individuals. These findings underline the importance of careful characterisation and modeling of exposure uncertainty in observational studies.

Introduction

Exposure measurement error is arguably one of the most important sources of uncertainty in epidemiological studies. It is widely acknowledged that when it is not or only poorly accounted for, measurement error can lead to biased risk estimates, a distortion of the shape of the exposure-response relationship and a loss in statistical power [1, 2]. Accounting for exposure measurement error can be daunting, however, because error characteristics tend to be complex in epidemiological studies.

In occupational cohort studies, for instance, one is usually interested in the association between the time until diagnosis or time until death by a certain disease and cumulative exposure to a certain chemical or physical agent. The analysis of this association may require the specification of a proportional hazards model where cumulative exposure is treated as a time-dependent variable. Owing to the time-dependent nature of cumulative exposure, the exposure history of a worker may be collected using different strategies according to the period of exposure. Changes in the methods of exposure assessment can create rather complex patterns of exposure uncertainty, where the type and magnitude of measurement error can vary over time. If no exposure data is available for the earliest years, one usually has to retrospectively reconstruct exposure values for this period. On the other hand, it is common to use a method of prospective, and possibly individual, exposure monitoring for the more recent exposure periods. In the periods of prospective exposure assessment, technical advances in measurement devices may imply more and more precise measures of exposure, which can translate into a decrease in measurement error over time. It has been suggested that the fact that exposure uncertainty and the magnitude of exposure are both highest for the earliest exposure periods may cause an attenuation of the exposure-response curve for high exposure values, a phenomenon frequently observed in occupational cohort studies [3–5].

As it is virtually impossible to reconstruct the exposure values for each individual worker in a retrospective fashion, one usually has to estimate the exposure levels for different job categories and the same exposure level is affected for all workers in a given job category. In this situation, individual exposure values of workers in a job-category are assumed to vary around the estimated exposure level and measurement error is therefore often described as unshared Berkson error [5–8], i.e., Berkson error that independently affects workers and different exposure values of the same worker. In this conception, the estimated exposure level is implicitly considered to be a precise estimate of the true average exposure in a job category, thereby neglecting the fact that many simplistic and potentially wrong assumptions typically have to be made in retrospective exposure reconstruction. Considering these uncertainties and simplifications, which often arise because working conditions may be very different from those in more recent years, the estimated exposure level can greatly differ from the true average exposure level in a job category. This discrepancy can be modelled as a classical measurement error component that is shared between workers. Indeed, it affects the exposure values of all workers in a given job category in the same way and therefore cannot be considered as independent for workers who belong to this job category. Several authors have described this error structure as a mixture of unshared Berkson and unshared classical measurement error [9, 10], but this view cannot account for the fact that the classical measurement error component affects all individuals in a group in the same way. At the same time, uncertainty components that are shared between individuals have received growing attention in the field of radiation epidemiology in recent years [11–15]. Quite contrary to [9, 10], the authors of [14, 15] account for the shared nature of error, but not for the fact that this shared error can be of Berkson or classical type [15, 16]. Meanwhile, comparatively little attention has been paid to the possibility of error components shared within workers in occupational cohort studies. When cumulative exposure is modelled as a time-dependent variable and a method of group-level exposure estimation is used, individual job conditions and worker practices may create a correlation between measurement errors in the exposure history of a worker [17, 18]. This correlation can be described by error shared within workers, i.e. an exposure uncertainty component shared for several years of the same worker. Through the summing of exposure values to obtain cumulative exposure, uncertainty components

shared within workers may be magnified, as the same error term is repeated for every exposure value of a worker. Therefore these components may have more impact on statistical inference than unshared uncertainty components or components that are shared between workers. On the other hand, several authors have argued that errors that are shared between individuals might have fundamentally different consequences on statistical inference than unshared measurement errors [13–16]. To our knowledge there are no studies confirming this assertion for proportional hazards models, which possibly presents the most widely applied class of models in medical research.

Stayner et al. [4] and Steenland et al. [5] examined the effects of heteroscedastic measurement error on the shape of the exposure-response curve and only found a modest attenuation of the exposure-response curve at high exposure values. However, the authors treated cumulative exposure in an occupational cohort as time fixed variable known at baseline, thereby ignoring both its time-varying nature and the possibility of exposure uncertainty components shared within individuals. In this context, further analyses are necessary to reassess these results when a more realistic structure of measurement error is assumed with the possibility of shared and unshared uncertainty components for all exposure periods.

The aim of this study is to highlight and compare the effects of shared and unshared exposure measurement error on risk estimation and the shape of the exposure-response relationship with the aid of simulated data when statistical inference in proportional hazards models is not corrected for exposure measurement error. In a first step, we will assume measurement models with only one type of error for all periods of exposure to compare the impact of different types of exposure uncertainty on risk estimation in two alternative proportional hazards models. In particular, we will investigate the influence of multiplicative Berkson and classical measurement error that can be shared between workers, within workers, both between workers and within workers, or unshared. In a second step, we will assume more complex measurement models with varying types of multiplicative measurement error for different exposure periods that reflect the conditions in an occupational cohort more realistically to assess the effects of these error structures on the shape of the exposure-response relationship.

Our motivating example concerns the potential impact of uncertainty in radon exposure when modelling lung cancer mortality in the French cohort of uranium miners.

Measurement models

In the following, we present several measurement models to describe shared and unshared exposure uncertainty components in an occupational cohort, which we will use in our simulation study. These measurement models describe the association between the true $X_{ij}(t)$ and the observed $Z_{ij}(t)$ exposure of worker i at time t, where worker i belongs to group j. A group can be formed by all workers belonging to a specific job-category in a retrospective exposure reconstruction or in a prospective method of group-level estimation. For the sake of simplicity, we will assume that each worker can only belong to one group and that the group he belongs to does not vary over time.

A large part of the measurement error literature is based on additive error. As it has been repeatedly suggested, however, that multiplicative measurement models may be more realistic in many situations in occupational and environmental epidemiology in general [6] and to describe uncertainty in airborne exposure in particular [3, 7, 19], we will assume a multiplicative log-normal model for exposure measurement error in the following. In order to describe exposure uncertainty in an occupational cohort, we will first consider simple measurement models in which the type (i.e., Berkson, classical, shared, unshared) and magnitude of error remains constant over the years. In other words, even though there may be several exposure periods, we will consider the same measurement error variance for all exposure periods. We will refer to these models as homoscedastic measurement models. In a second step, we will focus on more complex measurement models that may describe exposure uncertainty more adequately in an occupational cohort. Contrary to the homoscedastic measurement models that assume pure shared or unshared measurement error for all exposure periods, we will consider in these complex models that the type of error and the measurement error variances can vary for the different exposure periods. Additionally, we will consider that measurement error occurring in a given exposure period can be characterised by a combination of shared and unshared exposure uncertainty components. We will refer to these models as heteroscedastic measurement models.

Homoscedastic measurement models

Unshared measurement error

When modelling measurement error, one commonly distinguishes Berkson and classical measurement error. Unshared Berkson error is often presumed when a group-level method of exposure estimation is used. In this case, an observed or estimated exposure value is assigned to a group of workers and the true exposure of each worker is supposed to randomly deviate from this observed exposure value. The variability of true exposure is bigger than the variability of observed exposure and measurement error is independent of observed exposure. Multiplicative Berkson error can be expressed by model

$$\boldsymbol{\mathcal{M}_1}: X_{ij}(t) = Z_j(t) \cdot U_{ij}(t),$$

where $Z_{ij}(t) = Z_j(t)$ for all workers i of group j and $E(U_{ij}(t)|Z_j(t)) = 1$, which implies that $E(X_{ij}(t)|Z_j(t)) = Z_j(t)$.

When individual measurements are obtained through a measuring device, on the other hand, a classical measurement error model is assumed, where the observed exposure of worker i in group j $Z_{ij}(t)$ randomly deviates from his true exposure. In contrast to model $\boldsymbol{\mathcal{M}_1}$, the variability of observed exposure is bigger than the variability of true exposure and measurement error is independent of true exposure. Multiplicative classical measurement error can be expressed by model

$$\boldsymbol{\mathcal{M}_2}: Z_{ij}(t) = X_{ij}(t) \cdot U_{ij}(t),$$

where $E(U_{ij}(t)|X_{ij}(t)) = 1$ implying that $E(Z_{ij}(t)|X_{ij}(t)) = X_{ij}(t)$. Contrary to measurement model $\boldsymbol{\mathcal{M}_1}$, we do not dispose of $E(X_{ij}(t)|Z_{ij}(t))$ in measurement model $\boldsymbol{\mathcal{M}_2}$. A common way to correct for measurement error in this situation is to use regression calibration, where $E(X_{ij}(t)|Z_{ij}(t))$ is modelled as a function of $Z_{ij}(t)$ and the parameters of this function are estimated on a validation sample.

We will assume in both models that log transformed measurement errors $log(U_{ij}(t))$ are independent and normal random variables with mean $-\frac{\sigma^2}{2}$ and variance σ^2, i.e.,

$\log(U_{ij}(t)) \sim \mathcal{N}(-\frac{\sigma^2}{2}, \sigma^2)$ [1]. This parametrisation ensures that $E(U_{ij}(t)|Z_j(t)) = 1$ and $E(U_{ij}(t)|X_{ij}(t)) = 1$. In particular, $U_{ij}(t)$ and $U_{i'j}(t')$ are independent for $i \neq i'$ and $t \neq t'$ in both $\mathcal{M}_1$ and $\mathcal{M}_2$. Under this independence assumption, measurement error is considered as unshared.

Shared measurement error

To describe exposure uncertainty components that are shared between or within workers, one can adapt the unshared Berkson and classical measurement error model described previously by modifying the assumptions on the structure of measurement error $U_{ij}(t)$.

For instance, if we suppose that the measurement error term $U_{ij}(t)$ equals $U_j(t)$ for all workers i in group j, we can obtain a Berkson and a classical measurement error model in which the errors are shared for a group of workers. Indeed, in these models, the same error component is presumed at time t for all subjects i belonging to group j (hence the term "shared between workers") and $U_j(t)$ and $U_{j'}(t')$ are independent if $j \neq j'$ or $t \neq t'$. We will denote $\mathcal{M}_3$ and $\mathcal{M}_4$ the measurement models describing Berkson and classical measurement error shared between workers.

Similarly, to describe Berkson or classical measurement error that is shared for several years of the same worker, we assume in this case that the measurement error term $U_{ij}(t)$ neither depends on time t nor on group j: $U_{ij}(t) = U_i \ \forall j \forall t$. The same error component is supposed for all years of exposure of worker i and U_i and $U_{i'}$ are independent if $i \neq i'$ (hence the term "shared within workers"). We will denote $\boldsymbol{\mathcal{M}_5}$ and $\boldsymbol{\mathcal{M}_6}$ the measurement models describing Berkson and classical measurement error shared within workers.

Finally, an error component may simultaneously affect all exposure values received by the workers in a certain group during an exposure period of several years in the same way. This error structure is likely to occur when recent exposure conditions are extrapolated in order to reconstruct exposure values in the past. These extrapolations are typically made for a group of workers and for a period of several years at the same time, as it is difficult to make more precise estimates in a retrospective exposure reconstruction. In this situation, the measurement error induced by estimating the

[1] Note that if $\log(W)$ follows a normal distribution with mean μ and standard deviation σ, W follows a log-normal distribution with $E(W) = \exp(\mu + \frac{\sigma^2}{2})$ and variance $Var(W) = \big(\exp(\sigma^2) - 1\big)\exp(2\mu + \sigma^2)$.

mean exposure level for a group j affects both all workers in this group and all exposure values received by these workers. Therefore, the measurement error term $U_{ij}(t)$ neither depends on time t nor on subject i, but only on the group j a subject belongs to: $U_{ij}(t) = U_j \ \forall i \forall t$. In this case, U_j and $U_{j'}$ are independent if $j \neq j'$. We will denote $\boldsymbol{\mathcal{M}_7}$ and $\boldsymbol{\mathcal{M}_8}$ the measurement models describing Berkson and classical measurement error shared both within and between workers.

For a more detailed presentation of models $\boldsymbol{\mathcal{M}_3}$ to $\boldsymbol{\mathcal{M}_8}$, see S1 Appendix.

Heteroscedastic measurement models with three exposure periods

In the following, we will extend the measurement models presented so far by considering three exposure periods for which both the type and the magnitude of exposure uncertainty can vary and by allowing for combinations of shared and unshared measurement error in each exposure period. In this section, we denote $X_{ij}^q(t)$ and $Z_{ij}^q(t)$ the true and observed exposure for worker i belonging to group j at time t in exposure period q.

Unshared measurement error for different exposure periods

Motivated by the exposure conditions in the French cohort of uranium miners (see the motivating example described in the presentation of the simulation study for more details), we consider the following model to allow for possible changes in the method of exposure assessment:

$$\boldsymbol{\mathcal{M}_9}: \begin{cases} X_{ij}^1(t) = Z_j^1 \cdot U_{ij}^1(t) \\ X_{ij}^2(t) = Z_j^2(t) \cdot U_{ij}^2(t) \\ Z_{ij}^3(t) = X_{ij}^3(t) \cdot U_{ij}^3(t) \end{cases}$$

where $\log(U_{ij}^q(t))$ are independent and normal random variables with mean $-\frac{\sigma_q^2}{2}$ and variance σ_q^2. We suppose three distinct exposure periods, $q \in \{1, 2, 3\}$, with a retrospective exposure reconstruction for the first period, a prospective method of group-level exposure estimation for the second period and prospective and individual exposure assessment for the third period. A similar reasoning will apply to many

occupational cohort studies, as exposure values for the earliest years of exposure are often reconstructed retrospectively by extrapolating exposure conditions from more recent exposure periods, while methods of prospective exposure assessment are available for more recent years of exposure [3, 20, 21]. The models can also be easily extended to more than three exposure periods. By allowing $\sigma_q^2 \neq \sigma_{q'}^2$ for $q \neq q'$, model $\boldsymbol{\mathcal{M}_9}$ can describe the varying type and magnitude of error in the three exposure periods by a heteroscedastic and unshared error structure. The error occurring in the first two exposure periods is described as Berkson error with $E(U_{ij}^1(t)|Z_j^1) = 1$ and $E(U_{ij}^2(t)|Z_j^2(t)) = 1$, following the assumption that a method of group-level exposure estimation will lead to unshared Berkson error. The observed exposure Z_j^1 for group j in the first period does not depend on time t as only one value is estimated for all exposure years in that group in a retrospective exposure reconstruction. On the other hand, observed exposure $Z_j^2(t)$ for group j in the second period depends on time t, as exposure values are estimated by a prospective group-level exposure assessment. The individual and prospective method of exposure assessment in the third exposure period, on the other hand, is supposed to produce independent classical measurement error, which implies that $E(U_{ij}^3(t)|X_{ij}^3(t)) = 1$.

Accounting for the imprecision of the measurement device in group-level exposure estimation

When describing exposure uncertainty in an occupational cohort, it may be advisable to account for the imprecision of the measurement device which is used to estimate the average exposure for a group of workers in a group-level exposure assessment. In this vein, we can suppose an additional classical measurement error component, which is shared between and within workers for the for the first exposure period, as this period was characterised by a retrospective exposure reconstruction and the use of a single error prone exposure estimate for a group of workers and several years of exposure. For the second period, the imprecision of the measurement device that is used to obtain measurements for a group of workers in a prospective fashion will likely give rise to a component of classical measurement error that is shared between workers, but not for several years of the same worker. The combination of shared and unshared

measurement error in the first and second exposure period can be expressed by model

$$\boldsymbol{\mathcal{M}_{10}}: \begin{cases} X_{ij}^1(t) = \chi_j^1 \cdot U_{ij}^1(t) \\ Z_j^1 = \chi_j^1 \cdot U_j^{1*} \\ X_{ij}^2(t) = \mathcal{L}_j^2(t) \cdot U_{ij}^2(t) \\ Z_j^2(t) = \mathcal{L}_j^2(t) \cdot U_j^{2*}(t) \\ Z_{ij}^3(t) = X_{ij}^3(t) \cdot U_{ij}^3(t). \end{cases}$$

where χ_j^1 and $\mathcal{L}_j^2(t)$ are latent intermediate variables. χ_j^1 can be interpreted as the true average exposure value of group j in the retrospective exposure reconstruction in the first period. $\mathcal{L}_j^2(t)$ represents the true average exposure value of group j at time t in the prospective group-level exposure assessment during the second period. In model $\boldsymbol{\mathcal{M}_{10}}$, we assume $E(U_{ij}^1(t)|\chi_j^1)$, $E(U_j^{1*}|\chi_j^1)$, $E(U_{ij}^2(t)|\mathcal{L}_j^2(t))$, $E(U_j^{2*}(t)|\mathcal{L}_j^2(t))$ and $E(U_{ij}^3(t)|X_{ij}^3(t))$ all equal to one. Moreover, $\log(U_j^{1*}) \sim \mathcal{N}(-\frac{\sigma_{1*}^2}{2}, \sigma_{1*}^2)$ and $\log(U_j^{2*}(t)) \sim \mathcal{N}(-\frac{\sigma_{2*}^2}{2}, \sigma_{2*}^2)$. The error term U_j^{1*} is supposed to be shared both between and within subjects and therefore only depends on group j. The error term $U_j^{2*}(t)$, on the other hand, is shared between, but not within subjects and therefore depends on group j and time t but not on worker i.

Accounting for individual worker practices in the periods of group-level exposure estimation

Several authors have pointed out that workers can receive systematically higher or lower exposure values than the exposure level that is measured for their job category, although they work in the same environment and perform basically the same tasks [17, 18]. For instance, a comparison between a prospective method of group-level exposure assessment and individual exposure assessment in the French cohort of uranium miners revealed that individual cumulative radon exposure was substantially underestimated for some workers but not for others when exposure was assessed at the group-level [22]. A possible explanation for this finding is that some of the workers sought relief from the strong airstream produced by a ventilation system in their break hours, thereby exposing themselves to very high radon concentrations. The ventilation

system was installed in the mines as a measure of radiation protection and the access to areas where the airflow was too weak was formally forbidden. Miners who infringed this rule received systematically higher true exposure values than their estimated exposure level for all years of exposure, which were characterised by a method of group-level exposure assessment. To account for the effect of these individual worker characteristics and worker practices, it seems adequate to model exposure uncertainty as a combination of a component of unshared Berkson error and a component of Berkson error shared for several years of a worker when a method of group-based exposure assessment is used, expressed by model

$$\boldsymbol{\mathcal{M}_{11}}: \begin{cases} X^1_{ij}(t) = Z^1_j(t) \cdot U^{1*}_{ij}(t) \cdot U^{1*}_i \\ X^2_{ij}(t) = Z^2_j(t) \cdot U^{2*}_{ij}(t) \cdot U^{2*}_i \\ Z^3_{ij}(t) = X^3_{ij}(t) \cdot U^3_{ij}(t). \end{cases}$$

In this model, we assume $\log(U^{1*}_i) \sim \mathcal{N}(-\frac{\sigma^2_{1*}}{2}, \sigma^2_{1*})$ and $\log(U^{2*}_i) \sim \mathcal{N}(-\frac{\sigma^2_{2*}}{2}, \sigma^2_{2*})$ with $E(U^{1*}_i|Z^1_j) = 1$ and $E(U^{2*}_i|Z^2_j(t)) = 1$.

Simulation study

Motivating example

To mimic the exposure conditions of a "true" occupational cohort, we used information on annual radon exposure of the French cohort of uranium miners [23,24] as basis for all simulations. Radon is a noble and radioactive gas, resulting from the decay of uranium 238. As radon is considered to be the second most important cause of lung cancer after smoking, the main outcome of interest when it comes to radon exposure is lung cancer mortality. Estimated excess relative risk coefficients per 100 working level months[2] (WLM) vary between 0.8 and 4.2 in occupational cohorts of miners [25] when risk estimation is not corrected for measurement error.

The French cohort of uranium miners consists of 5086 uranium miners, who present

[2]Excess relative risk (ERR) is related to relative risk (RR) by the relation $RR = 1 + ERR$. Radon exposure in cohorts of underground miners is classically expressed in working level months with one working level month approximately equal to 6.3×10^5Bq h m^{-3}.

an average follow-up of 35 years between 1945 and 2007 and an average duration of employment of 17 years between 1945 and 1999. The radon exposure data of this cohort reflect conditions which can be seen as typical for an occupational cohort with varying methods of exposure assessment, depending on the period of exposure. For the earliest period of mining (1945-1955), there was no systematic radon exposure monitoring in the mines and exposure values had to be reconstructed retrospectively by a group of experts. The second exposure period (1956-1982) was characterised by a method of group-based exposure monitoring, where information gathered through ambient measurements at work sites was used to estimate the individual exposure of each miner in a prospective fashion. Finally, for the latest period of exposure (1983-1999), radon exposure was assessed individually and prospectively via personal dosimetry. At the same time, starting in 1955, improvements in radiation protection of the workers led to a sharp exposure reduction between 1955 and 1956, which was subsequently followed by a continual decrease in annual radon exposure until the last mine closed in France in 1999. The average radon exposure value for exposed miners was 28.28 WLM in 1955 and 0.14 WLM in 1999.

Despite evidence of multiple sources of shared uncertainty in radon exposure in cohorts of underground miners, attempts to model measurement error in these cohorts so far mainly relied on the hypothesis that all exposure uncertainty can be described by unshared multiplicative measurement error [7, 8, 26–28]. In the French cohort of uranium miners, an attenuation of the exposure response curve for cumulative exposure values exceeding 100 WLM has been observed [26]. This shape of the exposure response curve, which has been observed in many other occupational cohorts [3, 4], persists after unshared measurement error is accounted for. Stram et al. (1999) [1] observe a similar phenomenon when analysing lung cancer mortality in the Colorado Plateau Uranium Miners cohort, where the risk for radon exposures received at a high-dose rate is estimated to be smaller than the risk estimated for radon exposures received at a low dose-rate. The authors refer to this phenomenon, which is commonly observed in cohorts of uranium miners, as the inverse dose-rate effect. When reanalysing lung cancer mortality in the cohort with revised exposure estimates based on a model-based imputation scheme, the authors find that the inverse dose-rate effect is greatly diminished. In this context, it is important to ascertain whether the attenuation of the

exposure-response relationship observed in the French cohort of uranium miners could be due to components of shared exposure uncertainty that are not accounted for.

Simulation study 1: The impact of shared and unshared measurement error on risk estimation

Based on our motivating example, we performed a series of simulations to assess the impact of shared and unshared exposure uncertainty components on risk estimation in proportional hazard models when statistical inference is not corrected for measurement error.

Models used for data generation

To generate failure times, we considered two alternative proportional hazards models $\mathcal{D}_1$ and $\mathcal{D}_2$ to describe the association between instantaneous hazard rate of death by lung cancer of miner i at age t, $h_i(t)$ and his cumulative radon exposure. Both disease models specify $h_i(t)$ as a function of cumulative radon exposure in 100 WLM, $X_i^{\text{cum}}(t)$ of worker i until time t and the baseline hazard $h_0(t)$ of lung cancer mortality at age t. They are given by

$$\boldsymbol{\mathcal{D}_1} : h_i(t) = h_0(t)(1 + \beta X_i^{\text{cum}}(t))$$

and

$$\boldsymbol{\mathcal{D}_2} : h_i(t) = h_0(t) \exp(\beta X_i^{\text{cum}}(t)).$$

$\mathcal{D}_1$ represents an excess hazard ratio (EHR) model, which is commonly used to describe the association between cancer mortality and exposure to radon and to other sources of ionising radiation. $\mathcal{D}_2$, on the other hand, is the more classical form of the Cox proportional hazards model. Cumulative radon exposure $X_i^{\text{cum}}(t)$ is a time-varying variable as it represents the sum over all annual exposure values that worker i in group j received before time t: $X_i^{\text{cum}}(t) = \sum_{u \leq t} X_{ij}(u)$.

We considered different measurement models to describe the association between true $X_{ij}(t)$ and observed exposure $Z_{ij}(t)$ of worker i at time t belonging to group j. We

used models $\boldsymbol{\mathcal{M}_1}$ to $\boldsymbol{\mathcal{M}_8}$ presented earlier, which assume one exposure period and specify Berkson and classical measurement error that is either unshared, shared between subjects, shared within subjects or shared both within and between subjects.

Data Generation

To create unshared exposure uncertainty, measurement errors $U_{ij}(t)$ were sampled independently from a log-normal distribution for each miner i at every time t. For components of exposure uncertainty shared within workers, we generated only one log-normally distributed error U_i for each worker i, which was then used for all times t at which this worker was exposed to radon. Conversely, for exposure uncertainty shared between workers, we generated one log-normally distributed error term, $U_j(t)$ at each time t, which was affected to all workers i in group j. Finally, for an error component, that was both shared between and within workers, i.e., at the same time among a group of workers and for several years of the same worker, we generated one log-normally distributed error U_j for each group of workers, which was applied for all times at which a worker belonging to group j was exposed. When generating exposure data with Berkson error, we made the assumption that observed exposure $Z_{ij}(t)$ was equal to the observed exposure values in the French cohort of uranium miners and multiplied $Z_{ij}(t)$ by an error term to obtain true exposure $X_{ij}(t)$. Conversely, when generating exposure data with classical measurement error, we made the assumption that true exposure $X_{ij}(t)$ was equal to the observed exposure values in the cohort and multiplied $X_{ij}(t)$ by an error term to obtain observed exposure $Z_{ij}(t)$. Since the French cohort of uranium miners did not present a natural partition into groups of workers, we created homogeneous groups of workers via a hierarchical ascendant clustering algorithm after a multiple factor analysis based on some covariates concerning job characteristics: principal type of mine, principal location and principal type of job.

We adapted a method proposed by Henry (2014) [29] to generate failure times for a proportional hazards model with time-dependent covariates. To generate survival times, we used a piecewise constant model to specify baseline hazard $h_0(t)$ in both the EHR and the Cox proportional hazards model. All data generation was done in R (version 3.3.1).

We chose a true risk coefficient of $\beta = 5$ for the EHR model, which is in the same

order of magnitude as the risk coefficient estimated in the French cohort of uranium miners when restricting analyses to exposure periods characterised by a prospective method of exposure assessment [24, 26]. In order to achieve a comparable strength of the association between radon exposure and lung cancer mortality for the Cox model and the EHR model, we chose a risk coefficient of $\beta = 2$ for the Cox model. For the sake of completeness, we included results for the EHR model with $\beta = 2$ and for the Cox model with $\beta = 5$ in Table 4 and Table 5 in S2 Appendix.

We compared the impact of large and moderate measurement error, corresponding to values for the variance of measurement error of $\sigma_\epsilon^2 = 0.8$ and $\sigma_\epsilon^2 = 0.1$, respectively. These values were chosen in accordance with the characterisation of exposure uncertainty in the French cohort of uranium miners [27, 30].

The combination of all possibilities for the disease model, the measurement model, for the value of the true risk coefficient and for the measurement error variance resulted in $2 \times 8 \times 2 \times 2 = 64$ distinct simulation scenarios. Additionally, we compared the results with risk estimates when radon exposure was observed without measurement error, i.e. under measurement model $\mathcal{M}_0$, resulting in 68 simulation scenarios in total.

Assessment of risk estimates

For each scenario, inference was based on 100 simulated data sets. We conducted inference for disease model $\mathcal{D}_1$ for data sets that were generated according to the EHR model. Likewise, we conducted inference for the disease model $\mathcal{D}_2$ for data sets that were generated according to the Cox model. Observed exposure $Z_{ij}(t)$ was treated as an error-free surrogate of true exposure $X_{ij}(t)$ in the analysis of the association between exposure and disease outcome, i.e. measurement error was not accounted for in risk estimation. We used a Metropolis-Hastings algorithm developed and tested in Python version 2.7 for Bayesian inference. We chose a centred normal prior distribution with a large variance (1000) for the risk coefficient β, which was truncated to guarantee the positivity of $h_i(t)$ in the EHR model. After checking convergence, inference was based on 20.000 iterations after an initial burn-in phase of 10.000 iterations (thin=1).

For each scenario, we estimated:

1. An overall 95% credible interval ($CI_{95\%}$), which was obtained by combining the

chains of the 100 replicates for each scenario and by determining the 2.5 and 97.5 quantiles of the corresponding pooled chain.

2. The relative bias of a Bayesian point estimator $\hat{\beta}$, given by $\frac{(\hat{\beta}-\beta)}{\beta}$, where β is the risk coefficient which served to generate the data. We used the posterior median as Bayesian point estimate for the risk coefficient $\hat{\beta}$.

3. The coverage rate of 95% credible intervals, which was calculated by counting the proportion of the 100 replicates for which the 95% credible interval included the true value of the coefficient β

4. The statistical power, which was estimated by counting the proportion of replicates for which the 95% credible interval for β excluded 0.

Simulation study 2: The effects of measurement error characteristics on the shape of the exposure-response curve

We performed a second series of simulations to assess the effects of different error structures on the observed shape of the exposure-response curve when risk estimation is not corrected for measurement error.

Models used for data generation

We used models $\mathcal{D}_1$ (EHR model) and $\mathcal{D}_2$ (Cox model), described earlier to generate mortality data as a function of true cumulative radon exposure $X_i^{\text{cum}}(t)$.

We considered model $\boldsymbol{\mathcal{M}_0}$ with no measurement error, model $\boldsymbol{\mathcal{M}_1}$ describing unshared Berkson error with one exposure period and the more complex and more plausible models $\boldsymbol{\mathcal{M}_9}, \boldsymbol{\mathcal{M}_{10}}$ and $\boldsymbol{\mathcal{M}_{11}}$. $\boldsymbol{\mathcal{M}_9}$ only describes unshared error, thereby accounting for differences in type and magnitude of error occurring in the three exposure periods, while $\boldsymbol{\mathcal{M}_{10}}$ and $\boldsymbol{\mathcal{M}_{11}}$ also allow for shared components of exposure uncertainty due to the imprecision of the measurement device and individual worker practices, respectively.

Data generation

For model $\boldsymbol{\mathcal{M}_{10}}$, the latent intermediate variables χ_j^1 and $\mathcal{L}_j^2(t)$ were set to the exposure values observed in the French cohort of uranium miners and both observed exposure $Z_{ij}(t)$ and true exposure $X_{ij}(t)$ were obtained by multiplying these intermediate variables with shared and unshared measurement error, respectively. For all other models, Berkson and classical measurement error were generated according to the strategy described earlier.

We chose $\beta = 2$ for the Cox model and $\beta = 5$ for the EHR model. Results for the EHR model with $\beta = 2$ and for the Cox model with $\beta = 5$ can be found in Figs S1 and S2 and Table 6 of S2 Appendix.

In order to be able to assess the impact of the structure of measurement error, rather than the magnitude of error, we tried to keep the total magnitude of error constant for all models. Therefore, we generated data according to model $\boldsymbol{\mathcal{M}_9}$ with $\sigma_1^2 = 0.8$, $\sigma_2^2 = 0.15$ and $\sigma_3^2 = 0.01$, following Allodji et al. [27,30]. To obtain a global variance of log-transformed errors comparable to this scenario, we set σ^2 equal to 0.2 in the homoscedastic Berkson error model $\boldsymbol{\mathcal{M}_1}$. In accordance with the characterisation of exposure uncertainty made by Allodji et al. [30], the variance parameters in model $\boldsymbol{\mathcal{M}_{10}}$ accounting for the imprecision on the measurement device as shared source of uncertainty, were chosen as $\sigma_1^2 = 0.09, \sigma_2^2 = 0.03$ and $\sigma_3^2 = 0.01, \sigma_{1*}^2 = 0.81$ and $\sigma_{2*}^2 = 0.12$. For the variance parameters in model $\boldsymbol{\mathcal{M}_{11}}$, accounting for individual worker practices as shared source of uncertainty, we chose the same values for the variance parameters as for model $\boldsymbol{\mathcal{M}_{10}}$. In doing so, we are able to compare the effects of shared exposure uncertainty due to the imprecision on the measurement device in a group-level exposure estimation and due to individual worker practices for a given error variance.

Assessing the shape of the exposure-response curve

We conducted statistical inference for both models $\boldsymbol{\mathcal{D}_1}$ and $\boldsymbol{\mathcal{D}_2}$ for all data sets, regardless of the disease model that was used for data generation, to study the effects of different error structures on disease model choice when inference is not accounted for

measurement error. The Deviance Information Criterion[3] (DIC) was used to compare the competing disease models with smaller DIC values indicating a better fit to the data.

To investigate the possibility of measurement error to induce a non-linear exposure-response relationship, we estimated parameter values in an EHR ($\boldsymbol{\mathcal{D}_3}$) and a Cox model ($\boldsymbol{\mathcal{D}_4}$) based on natural cubic splines. In these models, we chose interior knots at the 20th, 40th, 60th and 80th percentile of the exposure distribution of cases, i.e. miners who died of lung cancer in our simulation study.

While these disease models allow for a graphical evaluation of the impact of different measurement error characteristics, the parameter estimates in these models are not easily interpretable. Consequently, we also fitted continuous piecewise-linear models with a breakpoint at 100 WLM to be able to complement the results of model $\boldsymbol{\mathcal{D}_3}$ and $\boldsymbol{\mathcal{D}_4}$ with slope estimates for high and low exposure values under the different error structures. $\boldsymbol{\mathcal{D}_3}$ and $\boldsymbol{\mathcal{D}_5}$ were estimated for data sets that were generated according to the linear EHR model $\boldsymbol{\mathcal{D}_1}$ to assess the effect of the different measurement error characteristics when the EHR model was the true disease model. Similarly, $\boldsymbol{\mathcal{D}_4}$ and $\boldsymbol{\mathcal{D}_6}$ were fitted for data sets that were generated according to the linear Cox model $\boldsymbol{\mathcal{D}_2}$. All statistical inference was based on the assumption that observed exposure $Z_{ij}^q(t)$ was a perfect surrogate of true exposure $X_{ij}^q(t)$. The natural cubic spline basis was constructed in R and Bayesian inference via a Metropolis-Hastings algorithm was conducted in Python. Inference was based on 20.000 iterations after an initial burn-in of 10.000 iterations (thin =1).

Results

The impact of shared and unshared measurement error on risk estimation

Table 1 shows risk estimates and overall 95% credible intervals in the Cox proportional hazards model. Exposure uncertainty shared within workers created more relative bias in risk estimates and smaller coverage rates than exposure uncertainty shared between workers. The relative bias of small measurement error of both Berkson and classical

[3]The DIC is a Bayesian model selection criterion, which can be seen as a penalised likelihood criterion evaluating the trade-off between goodness of fit and model complexity.

nature, for instance, was more than twice as big when this error was shared within workers rather than between workers. In general, the impact of unshared uncertainty and uncertainty shared between workers was comparable. Error components, which were both shared between and within individuals produced about as much bias as error components that were only shared within individuals.

Table 1. Average posterior median ($\hat{\beta}$), overall 95% credible intervals ($CI_{95\%}$), relative bias and coverage rate for 100 data sets generated according to the Cox model $\mathcal{D}_2$, a measurement model among $\mathcal{M}_0$ to $\mathcal{M}_8$ and a true risk coefficient of $\beta = 2$ per 100 WLM

Model	Type of sharing	Type of error	Error variance	$\hat{\beta}$	$CI_{95\%}$	Relative bias	Coverage rate
$\mathcal{M}_1$	unshared	Berkson	0.1	1.81	[1.64; 1.99]	-0.10	0.10
			0.8	1.25	[0.97; 1.49]	-0.38	0.00
$\mathcal{M}_2$		classical	0.1	1.75	[1.55; 1.93]	-0.13	0.02
			0.8	0.83	[0.45;1.21]	-0.59	0.00
$\mathcal{M}_3$	between	Berkson	0.1	1.82	[1.62; 2.01]	-0.09	0.22
			0.8	1.25	[1.03; 1.47]	-0.38	0.00
$\mathcal{M}_4$		classical	0.1	1.75	[1.53; 1.94]	-0.13	0.05
			0.8	0.80	[0.44; 1.16]	-0.60	0.00
$\mathcal{M}_5$	within	Berkson	0.1	1.45	[1.12, 1.69]	-0.28	0.00
			0.8	0.76	[0.54; 0.97]	-0.62	0.00
$\mathcal{M}_6$		classical	0.1	1.33	[1.04; 1.58]	-0.34	0.00
			0.8	0.39	[0.17; 0.62]	-0.81	0.00
$\mathcal{M}_7$	both	Berkson	0.1	1.46	[1.11; 1.78]	-0.27	0.00
			0.8	0.77	[0.54; 1.00]	-0.62	0.00
$\mathcal{M}_8$		classical	0.1	1.42	[1.04; 1.72]	-0.29	0.00
			0.8	0.49	[0.13; 0.86]	-0.76	0.00
$\mathcal{M}_0$	none	none	0	1.96	[1.80; 2.13]	-0.02	0.95

Table 2 presents the same summary statistics concerning risk estimates as Table 1 but for failure times generated according to the EHR model. The relative bias introduced by measurement error in the EHR model was smaller than the bias introduced in the Cox model. For big measurement error in the EHR model, we observed the same pattern as for the Cox model where measurement errors shared within workers caused more relative bias and lower coverage rates than unshared measurement error or measurement error that was only shared between workers. For small measurement error, this tendency was less evident. In general, classical

measurement caused more relative bias and smaller coverage rates than Berkson error and big measurement error caused more relative bias and smaller coverage rates than small measurement error, regardless of the disease model and regardless of whether exposure uncertainty was shared or unshared. When data were generated without exposure measurement error, the coverage rates of 95% credible intervals were very close to 95%.

The statistical power was estimated to be 100% for all scenarios, both for data generated according to the Cox model and according to the EHR model.

Table 2. Average posterior median ($\hat{\beta}$), overall 95% credible intervals ($CI_{95\%}$), relative bias and coverage rate for 100 data sets generated according to the EHR model $\mathcal{D}_1$, a measurement model among $\mathcal{M}_0$ to $\mathcal{M}_8$ and a true risk coefficient of $\beta = 5$ per 100 WLM

Model	Type of sharing	Type of error	Error variance	$\hat{\beta}$	$CI_{95\%}$	Relative bias	Coverage rate
$\mathcal{M}_1$	unshared	Berkson	0.1	4.87	[3.07; 7.68]	-0.03	0.93
			0.8	4.65	[2.89; 7.18]	-0.07	0.91
$\mathcal{M}_2$		classical	0.1	4.88	[3.13;7.47]	-0.02	0.94
			0.8	4.34	[2.71; 6.70]	-0.13	0.78
$\mathcal{M}_3$	between	Berkson	0.1	4.77	[3.14; 7.11]	-0.05	0.99
			0.8	4.69	[2.91; 7.31]	-0.06	0.93
$\mathcal{M}_4$		classical	0.1	4.79	[3.04; 7.35]	-0.04	0.93
			0.8	4.44	[2.82; 6.72]	-0.11	0.85
$\mathcal{M}_5$	within	Berkson	0.1	4.88	[3.13; 7.47]	-0.02	0.94
			0.8	3.98	[2.43; 6.23]	- 0.20	0.73
$\mathcal{M}_6$		classical	0.1	4.75	[3.01; 7.31]	-0.05	0.91
			0.8	3.03	[1.86; 4.71]	- 0.39	0.13
$\mathcal{M}_7$	both	Berkson	0.1	4.88	[3.11; 7.69]	-0.02	0.94
			0.8	3.86	[2.19; 6.59]	-0.23	0.55
$\mathcal{M}_8$		classical	0.1	4.76	[2.94; 7.29]	- 0.05	0.92
			0.8	3.15	[1.62; 5.25]	- 0.37	0.25
$\mathcal{M}_0$	none	none	0	4.90	[3.14; 7.45]	-0.02	0.96

The effects of measurement error characteristics on the shape of the exposure-response curve

As can be seen in Fig 1, exposure-response curves for data generated according to the Cox model with no measurement error ($\boldsymbol{\mathcal{M}_0}$) or unshared and homoscedastic Berkson error ($\boldsymbol{\mathcal{M}_1}$) were close to linear on the log-scale. Heteroscedastic unshared error ($\boldsymbol{\mathcal{M}_9}$) appeared to create a slightly non-linear association. Indeed, Table 3 confirms that the

Fig 1. Estimated exposure-response curve when fitting the Cox model $\boldsymbol{\mathcal{D}_4}$ based on natural cubic splines when data are generated according to the Cox model $\boldsymbol{\mathcal{D}_2}$ with a risk coefficient of $\boldsymbol{\beta = 2}$. (a) $\mathcal{M}_0$, i.e., no measurement error (b) $\mathcal{M}_1$, i.e., unshared and homoscedastic Berkson error, (c) $\mathcal{M}_9$, i.e., unshared error of Berkson and classical type (d) $\mathcal{M}_{10}$, i.e., heteroscedastic error with a shared classical component describing the imprecision of the measurement device and (e) $\mathcal{M}_{11}$, i.e., heteroscedastic error with a shared Berkson component describing individual worker practices

slope for exposure under 100 WLM in this scenario is estimated to be more than twice as big as for exposure values over 100 WLM. Moreover, according to the DIC, the EHR model fitted the data better than the Cox model in 34% of cases when exposure data were generated following this unshared and heteroscedastic measurement model, even though the true disease model was the Cox model. For data generated according to the measurement models which incorporated shared sources of uncertainty ($\boldsymbol{\mathcal{M}_{10}}$ and $\boldsymbol{\mathcal{M}_{11}}$) the attenuation of the exposure-response curve at high exposure values was even more noteworthy. The slope estimates for low exposures in these scenarios are about six to eight times bigger than the slope estimates for high exposures. Under these scenarios, DIC values indicated for all replicates that the EHR model fitted the data better than the Cox model, although data were generated according to the Cox model. In the three scenarios using heteroscedastic measurement models ($\boldsymbol{\mathcal{M}_9}, \boldsymbol{\mathcal{M}_{10}}$ and $\boldsymbol{\mathcal{M}_{11}}$), risk coefficients estimated in the piecewise-linear disease model $\boldsymbol{\mathcal{D}_6}$ were overestimated for low exposures and underestimated for high exposure values when data was generated according to the Cox model. Overall, we only observed a substantial attenuation of the exposure-response curve in the Cox model when the first exposure period was characterised by a mixture of unshared and shared measurement error, which was either shared within workers or both within and between workers.

Fig 2 suggests that the different patterns of shared or unshared measurement error

Table 3. Comparison of risk estimates when data are generated according to different disease and measurement models. $\text{DIC}_{EHR} < \text{DIC}_{Cox}$ gives the percentage of realisations for which the Deviance Information Criterion (DIC) was smaller for the Excess Hazard Ratio (EHR) model when the true model was the Cox model and vice versa for $\text{DIC}_{Cox} < \text{DIC}_{EHR}$. The difference in DIC is calculated as difference between the EHR model and the Cox model.

Disease model	**Model $\mathcal{M}_0$** No error	**Model $\mathcal{M}_1$** Unshared Berkson error	**Model $\mathcal{M}_9$** Unshared heteroscedastic Berkson and classical error	**Model $\mathcal{M}_{10}$** Heteroscedastic shared device	**Model $\mathcal{M}_{11}$** Heteroscedastic worker practices
Data generated according to the Cox model ($\mathcal{D}_2$) with $\beta = 2$					
Risk estimate $\hat{\beta}$ in the linear Cox model ($\mathcal{D}_2$)	1.97 [1.78; 2.16]	1.67 [1.50; 1.87]	1.23 [1.00; 1.42]	0.57 [0.21; 1.06]	0.77 [0.59; 0.98]
Risk estimates in the piecewise-linear Cox model ($\mathcal{D}_6$)					
$\hat{\beta}_1$ (under 100 WLM)	1.98 [1.57; 2.40]	2.08 [1.65; 2.49]	2.21 [1.78; 2.61]	2.50 [2.06; 2.91]	2.33 [1.93; 2.70]
$\hat{\beta}_2$ (over 100 WLM)	1.96 [1.68; 2.26]	1.49 [1.22; 1.80]	0.92 [0.64; 1.18]	0.31 [0.06; 0.68]	0.40 [0.20; 0.63]
$\text{DIC}_{EHR} < \text{DIC}_{Cox}$	0%	0%	34 %	99%	100%
Difference in DIC	-216.08	-142.13	-15.17	169.16	107.24
Data generated according to the EHR model ($\mathcal{D}_1$) with $\beta = 5$					
Risk estimate $\hat{\beta}$ in the in the linear EHR model ($\mathcal{D}_1$)	4.90 [3.24; 7.62]	4.71 [3.08; 7.19]	4.44 [2.93; 6.81]	4.11 [2.26; 7.21]	4.07 [2.49; 6.28]
Risk estimates in the piecewise-linear EHR model ($\mathcal{D}_5$)					
$\hat{\beta}_1$ (under 100 WLM)	4.95 [2.83; 8.33]	4.81 [2.91; 7.67]	4.75 [2.79; 7.59]	5.58 [3.38; 9.16]	4.73 [2.77; 7.64]
$\hat{\beta}_2$ (over 100 WLM)	5.14 [2.06; 9.17]	4.72 [2.05; 9.21]	4.16 [1.48; 7.71]	2.18 [0.27; 6.43]	3.09 [0.69; 6.40]
$\text{DIC}_{Cox} < \text{DIC}_{EHR}$	0%	0%	0%	0%	0%
Difference in DIC	93.76	87.98	85.62	132.64	83.10

Fig 2. Estimated exposure-response curve when fitting the Excess Hazard Ratio (EHR) model $\mathcal{D}_3$ based on natural cubic splines when data are generated according to the EHR model $\mathcal{D}_1$ with a risk coefficient of $\beta = 5$. (a) $\mathcal{M}_0$, i.e., no measurement error (b) $\mathcal{M}_1$, i.e., unshared and homoscedastic Berkson error, (c) $\mathcal{M}_9$, i.e., unshared error of Berkson and classical type (d) $\mathcal{M}_{10}$, i.e., heteroscedastic error with a shared classical component describing the imprecision of the measurement device and (e) $\mathcal{M}_{11}$, i.e., heteroscedastic error with a shared Berkson component describing individual worker practices

did not produce any notable attenuation in exposure-response curves when mortality data were generated according to the EHR model. The risk estimates in the piecewise linear EHR model in Table 3 reveal that the risk for exposures under 100 WLM is estimated to be more than twice as big as the risk estimated for exposures exceeding 100 WLM when exposure data is contaminated with components of Berkson error which are shared for several years of the same worker ($\boldsymbol{\mathcal{M}_{10}}$). This measurement model is the only model for which the risk estimate for low exposures is estimated to be higher than the risk coefficient that was chosen to generate the data. For shared error components reflecting the imprecision of the measurement device ($\boldsymbol{\mathcal{M}_{11}}$), we also observe a higher risk estimate for exposures under 100 WLM than for exposures exceeding 100 WLM in the piecewise linear EHR model. However, in both cases, the credible intervals for the parameters in the piecewise linear model are very large and overlap. In contrast to the Cox model, DIC values always indicated the EHR to be the better fitting disease model when failure times were generated according to the EHR model, regardless of the measurement model.

Discussion

In the present simulation study, we compared the effects of shared and unshared uncertainty in cumulative exposure in an occupational cohort study on risk estimation and on the shape of the exposure-response relationship in proportional hazards models. In general, exposure uncertainty shared within individuals (i.e., shared for several years of exposure for an individual) caused more bias in risk estimates and smaller coverage rates than unshared exposure uncertainty. In contrast to claims that uncertainty shared between individuals should have fundamentally different effects on parameter estimation than unshared exposure uncertainty [15, 16], we found that both error components resulted in comparable relative bias and coverage rates in risk estimation in proportional hazard models. In line with previous findings on the impact of measurement error, we found that classical measurement error had more impact on inference than Berkson error [2], regardless of the extent and type of sharing. While we chose a Bayesian approach to conduct statistical inference on risk estimates, frequentist likelihood-based inference yields the same results when it comes to the bias introduced

by different components of measurement error (results not shown) as we assumed flat prior distributions. In line with this argument, concerning the relative bias in risk estimates in the presence of large and moderate unshared Berkson error, we observed values that were consistent with the results of Bender et al. (2005) and Küchenhoff et al. (2007), who studied the effect of unshared additive and multiplicative Berkson error on frequentist inference conducted for the Cox model. When studying the association between a disease outcome and cumulative exposure, we found that measurement error shared within individuals had more impact on risk estimation than measurement error shared between individuals. This finding is in accordance with the general principle that the impact of measurement error strongly depends on the variance of exposure and the variance of measurement error [2, 31]. In order to obtain cumulative exposure values in an occupational cohort study, the annual exposure values for a worker have to be summed and an error term shared within workers will be repeated for several exposure values in that sum. As the variance of the sum of positively correlated variables is greater than the sum of their variances, this summing will increase the measurement error variance in cumulative exposure. Uncertainty components shared between workers are unlikely to have a similar effect, because exposure values are summed within workers and not between workers. While it is therefore not surprising that error components shared within workers have more impact on statistical inference than components shared between workers when the main risk factor of interest is cumulative exposure, this result has important implications for the analysis of occupational cohort studies. In particular, this finding casts doubt on the common practice to model measurement error occurring in the exposure history of a worker on the sum of these values [4, 5, 32], instead of modelling on their natural level of occurrence, namely on the monthly or annual exposure values. In making this simplifying assumption, one may mistakenly model an error component that is shared for several years of a worker as an unshared error component. Our results suggest that may yield highly misleading results.. We found that the impact of error components that were shared both within and between workers was comparable to the impact of error components that were shared between workers. However, it is likely that an error component that is shared for all members of a cohort could have even larger effects on statistical inference than an error component that is only shared for a sub-group of workers.

In accordance with the results obtained by Steenland et al. [5], we only observed a mild attenuation of the exposure-response curve in the Cox model when assuming a structure of unshared error in which the magnitude of error and the magnitude of exposure was greatest for the earliest years of exposure, which are often characterised by retrospective exposure reconstruction. However, in an occupational cohort, it seems more plausible to assume shared error components due to the imprecision of the measurement device and individual worker practices when exposure values are retrospectively reconstructed. Under these assumptions, we found a considerable attenuation in the exposure-response relationship for high exposure values when data were generated according to the Cox model. Attenuations of the exposure-response curve at high exposure values may pose serious challenges in risk modelling in occupational cohort studies. Indeed, if this attenuation reflects the association between true exposure and the outcome and a linear model is chosen, it may cause a severe underestimation of risk for workers with low exposures. On the other hand, if the association between true exposure and the outcome is linear and the observed distortion of the exposure-response relationship is caused by measurement error, fitting a non-linear or a piecewise-linear model can lead to an overestimation of the risk coefficient for workers with low exposures. To support radiation protection, researchers are particularly interested in the low exposure range, because exposure levels of workers are currently much lower than in the past. Moreover, these exposure values are comparable to exposures received by the general population. Ignoring the cause of an observed distortion of the exposure-response curve may therefore seriously limit the extrapolability of risk estimates obtained in occupational studies to the general population.

In accordance with previous findings concerning the relative importance of measurement error in linear and log-linear models [32], we found that distortions in the exposure-response relationship were more severe when data were generated according to the Cox model, rather than according to the EHR model. Moreover, when failure times were generated according to the Cox model and observed exposure values were contaminated with shared and unshared error, DIC values identified the EHR model as the model that best fitted the data. On the one hand, the robustness to measurement error makes the EHR model, which is often considered the best model to describe the

effects of ionising radiation on mortality, attractive for risk modelling in epidemiological studies. On the other hand, this finding casts doubt on the possibility to identify a "true model" to describe the exposure-risk relationship when risk estimates are not corrected for all sources of exposure uncertainty.

Concerning the impact of measurement error in radon exposure in the French cohort of uranium miners, our findings strengthen the hypothesis that the observed attenuation of the exposure response relationship might be caused by components of shared measurement error, as these components are likely to have occurred in the first exposure period of the cohort. Moreover, they call into question the results of previous studies accounting for exposure uncertainty, as these studies relied on the hypothesis that all exposure uncertainty occurring in this cohort could be described by unshared measurement error [26–28].

More generally, the results of the present study underline the importance of making a careful characterisation of shared and unshared exposure uncertainty in observational studies if the aim is to account for its potential impacts on statistical inference. In particular, one should be aware of the distortions of the exposure response relationship that may be induced by different degrees of precision and varying amounts of sharing. To obtain corrected risk estimates, it is important to use statistical methods that allow for complex patterns of shared and unshared measurement error. As measurement error shared within individuals appears to have more impact on risk estimation than unshared error components or error components shared between individuals, it is important to correctly specify these error components as such and to account for the fact that the type of exposure uncertainty may vary over time. Up to our knowledge, there is currently no possibility to use classical methods, such as regression calibration or simulation extrapolation to handle these complex patterns of measurement error. In our view, the Bayesian hierarchical approach is the most promising framework in this context [26, 33]. It is arguably the most flexible approach to account for exposure uncertainty and corrected parameter estimates can be obtained by Markov Chain Monte Carlo sampling. Additionally, the integration of prior knowledge on unknown parameters available from previous studies or in the form of expert knowledge can lead to more precise risk estimates and help to avoid overfitting, thereby increasing the replicability of findings.

The results of the present study may not only provide new insights in the interpretation and the discussion of analyses conducted on current occupational cohorts, but also for the design of future epidemiological studies. Methods of individual exposure assessment are becoming more accessible than ever with technical advances that facilitate the collection of exposure data. It is often argued that exposure uncertainty in group-level exposure estimation will not bias risk estimates, by combining the two simplifying assumptions that a group-level exposure estimation leads to Berkson error and that Berkson error does not bias risk estimates [5, 6, 34]. The results of the present study suggest that both of these simplifying assumptions do not hold in general and that shared components of Berkson error can even lead to a substantial distortion of the exposure-response relationship in the Cox model. In our view, a method of individual exposure assessment should be preferred over a method of group-level exposure estimation to avoid uncertainty components shared within workers and between workers, which may arise in a method of group-level exposure estimation because of the imprecision of the measurement device and individual worker practices.

Supporting information

S1 Fig. Estimated exposure-response curve when fitting the Cox model $\mathcal{D}_4$ based on natural cubic splines when data are generated according to the Cox model $\mathcal{D}_2$ with a risk coefficient of $\beta = 2$. (a) $\mathcal{M}_0$, i.e., no measurement error (b) $\mathcal{M}_1$, i.e., unshared and homoscedastic Berkson error, (c) $\mathcal{M}_9$, i.e., unshared error of Berkson and classical type (d) $\mathcal{M}_{10}$, i.e., heteroscedastic error with a shared classical component describing the imprecision of the measurement device and (e) $\mathcal{M}_{11}$, i.e., heteroscedastic error with a shared Berkson component describing individual worker practices

S2 Fig. Estimated exposure-response curve when fitting the Excess Hazard Ratio (EHR) model $\mathcal{D}_3$ based on natural cubic splines when data are generated according to the EHR model $\mathcal{D}_1$ with a risk coefficient of $\beta = 5$. (a) $\mathcal{M}_0$, i.e., no measurement error (b) $\mathcal{M}_1$, i.e., unshared and homoscedastic Berkson error, (c) $\mathcal{M}_9$, i.e., unshared error of Berkson and classical type (d) $\mathcal{M}_{10}$, i.e.,

heteroscedastic error with a shared classical component describing the imprecision of the measurement device and (e) $\mathcal{M}_{11}$, i.e., heteroscedastic error with a shared Berkson component describing individual worker practices

S3 Fig. Estimated exposure-response curve when fitting the Excess Hazard Ratio (EHR) model $\mathcal{D}_3$ based on natural cubic splines when data are generated according to the EHR model $\mathcal{D}_1$ with a risk coefficient of $\beta = 5$ assuming additive measurement error (a) $\mathcal{M}_0$, i.e., no measurement error (b) $\mathcal{M}_1$, i.e., unshared and homoscedastic Berkson error, (c) $\mathcal{M}_9$, i.e., unshared error of Berkson and classical type (d) $\mathcal{M}_{10}$, i.e., heteroscedastic error with a shared classical component describing the imprecision of the measurement device and (e) $\mathcal{M}_{11}$, i.e., heteroscedastic error with a shared Berkson component describing individual worker practices

S1 File. A more detailed presentation of measurement models $\mathcal{M}_5$, $\mathcal{M}_6$, $\mathcal{M}_7$ and $\mathcal{M}_8$.

S2 File. Results for alternative values of risk coefficients and concerning the attenuation of the exposure-response relationship introduced by additive measurement error

Acknowledgments

This work was partially supported by AREVA NC, in the framework of a bilateral agreement between IRSN and AREVA NC. AREVA NC had no role in study design, data analysis, or in the interpretation of the results.

References

1. **Stram D, Langholz B, Huberman M, Thomas D. Correcting for exposure measurement error in a reanalysis of lung cancer mortality for the Colorad Plateau uranium miners cohort. Health physics. 1999;77(3).**

2. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. Measurement error in nonlinear models: a modern perspective. Boca Raton: Chapman Hall; 2006.

3. Hertz-Picciotto I, Smith AH. Observations on the dose-response curve for arsenic exposure and lung cancer. Scandinavian Journal of Work, Environment & Health. 1993;19:217–226.

4. Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scandinavian Journal of Work, Environment & Health. 2003;29:317–324.

5. Steenland K, Karnes C, Darrow L, Barry V. Attenuation of exposure-response rate ratios at higher exposures: A simulation study focusing on frailty and measurement error. Epidemiology. 2015;26(3):395–401.

6. Armstrong BG. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occupational and Environmental Medicine. 1998;55(10):651–656.

7. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. Statistics in Medicine. 2005;24(11):1713 – 1723.

8. Küchenhoff H, Bender R, Langner I. Effect of Berkson measurement error on parameter estimates in Cox regression models. Lifetime Data Analysis. 2007;13(2):261–272.

9. Reeves GK, Cox DR, Darby SC, Whitley E. Some aspects of measurement error in explanatory variables for continuous and binary regression models. Statistics in Medicine. 1998;17:2157 –2177.

10. Mallick B, Hoffman FO, Carroll RJ. Semiparametric regression modeling with mixtures of Berkson and classical error, with

application to fallout from the Nevada test site. Biometrics. 2002;58(1):13 – 20.

11. Stram DO, Kopecky KJ. Power and uncertainty analysis of epidemiological studies of radiation-related disease risk in which dose estimates are based on a complex dosimetry system: Some observations. Radiation Research. 2003;160(4):408–417.

12. Little MP, Kukush AG, Masiuk SV, Shkylar S, Carroll RJ, Lubin JH, et al. Impact of uncertainties in expsoure assessment on estimates of thyroid cancer risk among Ukranian children and adolescents exposed from the Chernobyl accident. PLoS one. 2014;9(1).

13. Stram DO, Preston DL, Sokolnikov M, Napier B, Kopecky KJ, Boice J, et al. Shared dosimetry error in epidemiological dose-response analyses. PLoS One. 2015;10(3):e0119418. doi:10.1371/journal.pone.0119418.

14. Zhang Z, Preston DL, Sokolnikov M, Napier BA, Degteva M, Moroz B, et al. Correction of confidence intervals in excess relative risk models using Monte Carlo dosimetry systems with shared errors. PLoS One. 2017;12(4).

15. Simon SL, Hoffman FO, Hofer E. The two-dimensional Monte Carlo: a new methodological paradigm for dose reconstruction for epidemiological research. Radiation Research. 2015;183:27–41.

16. Kwon D, Hoffman FO, Moroz BE, Simon SL. Bayesian dose-response analysis for epidemiological studies with complex uncertainty in dose estimation. Statistics in medicine. 2016;35(3):399–423.

17. Greenland S, Fischer HJ, Kheifets L. Methods to Explore Uncertainty and Bias Introduced by Job Exposure Matrices. Risk Analysis. 2016;36(1):74–82.

18. Kromhout H. Design of measurement strategies for workplace exposures. Occupational and Environmental Medicine. 2002;59:349 – 354.

19. Lyles R, Kupper L. A detailed evaluation of adjustement methods for multiplicative measurement error in linear regression with applications in occupational epidemiology. Biometrics. 1997;53(3):1008–1025.

20. Gibb HJ, Lees PS, Pinsky PF, Rooney BC. Lung cancer among workers in chromium chemical production. American Journal of Industrial Medicine. 2000;38(2):115–126.

21. Cocco P, Rice CH, Chen JQ, McCawley MA, McLaughlin JK, Dosemeci M. Lung cancer risk, silica exposure, and silicosis in Chinese mines and pottery factories: the modifying role of other workplace lung carcinogens. American Journal of Industrial Medicine. 2001;40(6):674–682.

22. Zettwoog P. State-of-the-art of the *alpha* individual dosimetry in France. In: Gomez M, editor. Radiation hazards in mining: Control, measurements and medical aspects; 1981. p. 4–9.

23. Vacquier B, Caer S, Rogel A, Feurprier M, Tirmarche M, Luccioni C, et al. Mortality risk in the French cohort of uranium miners: extended follow-up 1946–1999. Occupational and Environmental Medicine. 2008;65(9):597–604.

24. Rage E, Caër-Lorho S, Drubay D, Ancelet S, Laroche P, Laurier D. Mortality analysis in the updated French cohort of uranium miners (1946 – 2007). International Archives of Occupational and Environmental Health. 2015;88(6):717–730.

25. Kreuzer M, Fenske N, Schnelzer M, Walsh L. Lung cancer risk at low radon exposure rates in German uranium miners. British Journal of Cancer. 2015;113(9):1367–1369.

26. Hoffmann S, Rage E, Laurier D, Laroche P, Guihenneuc C, Ancelet S. Accounting for Berkson and classical measurement error in radon exposure using a Bayesian structural approach in the analysis of lung cancer mortality in the French cohort of uranium miners. Radiation Research. 2017;187(2):196–209.

27. Allodji RS, Leuraud K, Thiébaut AC, Henry S, Laurier D, Bénichou J. Impact of measurement error in radon exposure on the estimated excess relative risk of lung cancer death in a simulated study based on the French Uranium Miners' Cohort. Radiation and Environmental Biophysics. 2012;51(2):151–163.

28. Allodji RS, Thiébaut A, Leuraud K, Rage E, Henry S, Laurier D, et al. The performance of functional methods for correcting non-Gaussian measurement error within Poisson regression: corrected excess risk of lung cancer mortality in relation to radon exposure among French uranium miners. Statistics in Medicine. 2012;31(30):4428–4443.

29. Hendry DJ. Data generation for the Cox proportional hazards model with time-dependent covariates: a method for medical researchers. Statistics in Medicine. 2014;33:436–454.

30. Allodji RS, Leuraud K, Bernhard S, Henry S, Bénichou J, Laurier D. Assessment of uncertainty associated with measuring exposure to radon and decay products in the French uranium miners cohort. Journal of Radiological Protection. 2012;32(1):85–100.

31. Heid I, Küchenhoff H, Wellmann J, Gerken M, Kreienbrock L, Wichmann HE. On the potential of measurement error to induce differential bias on odds ratio estimates: an example from radon epidemiology. Statistics in Medicine. 2002;21:3261–3278.

32. Steenland K, Deddens J, Zhao S. Biases in estimating the effect of cumulative exposure in log-linear models when estimated exposure

levels are assigned. Scandinavian Journal of Work, Environment & Health. 2000;26(1):37–43.

33. Bartlett JW, Keogh RH. Bayesian correction for covariate measurement error: A frequentist evaluation and comparison with regression calibration. Statistical Methods in Medical Research. 2016;.

34. Tielemans E, Kupper L, Kromhout H, Heederik D, Houba R. Individual-based and group-based occupational exposure assessment: Some equations to evaluate different strategies. Annals of occupational Hygiene. 1998;42(2):115–119.

Noname manuscript No.
(will be inserted by the editor)

A cautionary comment on the generation of Berkson error in epidemiological studies

Sabine Hoffmann · Chantal Guihenneuc · Sophie Ancelet

the date of receipt and acceptance should be inserted later

Abstract Exposure measurement error can be seen as one of the most important sources of uncertainty in studies in epidemiology. It is indispensable to dispose of reliable methods for measurement error generation when the aim is to assess the effects of measurement error or to compare the performance of several methods for measurement error correction. This paper compares two approaches for the generation of Berkson error, which have recently been applied in radiation epidemiology, in their ability to generate exposure data that satisfy the properties of the Berkson model. In particular, we show that the use of one of the methods produces results that are not in accordance with two important properties of Berkson error.

Keywords Radon · Measurement error · Uranium miners

1 Introduction

Exposure measurement error is unavoidable in most epidemiological studies. In cases where it is not or only poorly accounted for, it can lead to bias in risk estimates, a distortion of the exposure-response relationship and a loss in statistical power [1]. It can therefore be seen as one of the most important sources of uncertainty in epidemiological studies [2]. In radiation epidemiology, we are presented with a further challenge, because disease outcomes may not be directly associated with exposure, but rather with radiation dose. Radiation dose does not only depend on the exposure to ionising radiation, but also on other uncertain input parameters in dose calculation. Therefore, uncertainty in radiation doses does not only arise because of exposure measurement error, but also because there is often a lack of knowledge on the exact exposure conditions. For the analysis and the interpretation of epidemiological findings, it is important to be able to assess the effects of exposure measurement error on risk estimation and to account for them. Several methods for the correction of measurement error rely on the generation of error-prone exposure data [3,4]. In particular, a number of methods for the correction of measurement error have been proposed in radiation epidemiology that are based on the generation of possible dose vectors reflecting complex uncertainty structures [3,5,6]. Moreover, one generally has to generate error-prone exposure data in simulation studies that aim to assess the effects of different types of measurement error on risk estimation and in studies that compare the performance of several methods for measurement error correction. The correct generation of exposure measurement error is therefore a crucial part in its treatment. One traditionally distinguishes classical measurement error and Berkson error. Allodji et al. [7,8] and Hoffmann et al. [9] employed two different approaches for the generation of Berkson error when treating exposure uncertainty in the French cohort of uranium miners. The aim of this paper is to compare these two simulation approaches in their ability to generate exposure data that satisfy the properties of the Berkson model.

S. Hoffmann · S. Ancelet
Institut de Radioprotection et de Sûreté Nucléaire (IRSN),
PSE-SANTE/SESANE/LEPID
BP 17
92262 Fontenay-aux-Roses
Tel.: +33-158357342
Fax: +33-146570386
E-mail: sabine.hoffmann@irsn.fr

C. Guihenneuc
Université Paris Descartes, Faculté de Pharmacie

2 Methods

When modeling exposure measurement error, one typically distinguishes the true X_i and the observed Z_i exposure of study participant $i, i = 1, \dots n$. Depending on the error structure, the measurement error term U_i will describe the difference or the ratio of true and observed exposure for additive and multiplicative error, respectively. While it common to suppose that measurement error U_i follows a normal distribution in the case of additive measurement error, it is convenient to suppose a lognormal distribution for multiplicative exposure measurement error. Another important dimension

Table 1 The Berkson and classical measurement error model in the additive and the multiplicative case

	Additive error	Multiplicative error
Berkson error	$X_i = Z_i + U_i$ $E(U_i\|Z_i) = 0$	$X_i = Z_i \cdot U_i$ $E(U_i\|Z_i) = 1$
Classical measurement error	$Z_i = X_i + U_i$ $E(U_i\|X_i) = 0$	$Z_i = X_i \cdot U_i$ $E(U_i\|X_i) = 1$

to describe the characteristics of measurement error is the distinction of Berkson and classical measurement error. In a Berkson model, true exposure X_i is modeled conditional on observed exposure Z_i, contrary to a classical measurement model, where observed exposure Z_i is modeled conditional on true exposure X_i. Moreover, under the assumption of normal additive or lognormal multiplicative error, the variability of true exposure exceeds the variability of observed exposure in the Berkson model, while the opposite is true for classical measurement error (property 1). Moreover, the measurement error term U_i is independent of observed exposure Z_i in the former and of true exposure X_i in the latter. We will consider the four measurement models summarized in Table 1.
Another important difference between these two measurement models concerns their impact on risk estimation. In particular, in the case of simple linear regression, analytical results show that classical measurement error will lead to an underestimation of the regression coefficient, while Berkson error does not bias the estimation of the risk coefficient in this model [1] (property 2).
In the following, we briefly describe the two different simulation approaches for the generation of Berkson error used in Hoffmann et al [9] and Allodji et al. [7,8] .

Method 1 (Hoffmann et al (2017) [9]): For multiplicative error, given observed exposure Z_i of study participant i, multiply this observed exposure value by a measurement error term U_i to generate true exposure X_i:

$$X_i = Z_i \cdot U_i \tag{1}$$

where U_i follows a log-normal distribution with parameters $\mu = -\frac{\sigma^2}{2}$ and σ. As the expectation of a log-normal distribution with mean μ and standard deviation σ is given by $\exp(\mu + \frac{\sigma^2}{2})$, this parameterization implies $E(U_i) = 1$. Moreover, since U_i is generated independently of Z_i, $E(U_i|Z_i) = E(U_i) = 1$ and $E(X_i|Z_i) = Z_i$, i.e. there is no systematic bias.
Likewise, additive Berkson error can be generated by $X_i = Z_i + U_i$, where U_i follows a normal distribution with mean zero and standard deviation σ.

Method 2 (Allodji et al (2012) [7,8]): For multiplicative error, which satisfies the Berkson model $X_i = Z_i \cdot U_i$ and given true exposure X_i of study participant i, divide this true exposure value by U_i to generate observed exposure Z_i:

$$Z_i = \frac{X_i}{U_i} \tag{2}$$

where U_i follows a log-normal distribution with parameters $\mu = -\frac{\sigma^2}{2}$ and σ.
Similarly, to obtain additive error, which satisfies the Berkson model $X_i = Z_i + U_i$, generate Z_i as $Z_i = X_i - U_i$, where U_i follows a normal distribution with mean zero and standard deviation σ.

In order to assess whether the exposure data generated by the two methods satisfied the properties of Berkson error, we generated 10000 datasets of 5000 workers. For each dataset, 5000 values of the measurement error term U_i ($i = 1, \dots, 5000$) were simulated according to a lognormal distribution with parameters $\mu = -\frac{\sigma^2}{2}$ and σ for each method. We considered two alternative values, 0.3 and 0.8, for the standard deviation parameter σ. Observed Z_i and true exposure X_i were obtained according to the two methods described above. Finally, we simulated 5000 outcome values Y_i following a simple linear regression model $Y_i = X_i\beta + \epsilon_i$ for each method, where β was equal to one and ϵ_i was identically and independently distributed following a normal distribution with mean zero and a standard deviation of one. To assess the impact of Berkson error on the estimation of β, we considered the following estimation model: $Y_i = Z_i\beta^* + \epsilon_i$, where Z_i is error-prone observed exposure. Maximum likelihood estimates of β^*, denoted $\hat{\beta}^*$, were computed via the lm function available in R version 3.1.1.

3 Results

Table 2 and Table 3 show the mean bias of $\hat{\beta}^*$, obtained by averaging the 10000 estimates $\hat{\beta}^*$ and subtracting the true value β, and the mean empirical variance of X_i and Z_i for each method for a standard deviation of 0.3 and 0.8, respectively. Moreover, the two tables give information on the 95% quantile interval of bias, defined as the empirical 2.5% and 97.5% quantiles of the 10000 computed values for bias.

Table 2 Mean bias, associated 95% quantile interval (95% QI) and mean empirical variance of the true exposure vector X and the observed exposure vector Z for Berkson error generated with a standard deviation of 0.3 and a true regression coefficient β of 1.

	Additive error		Multiplicative error	
	Method 1	Method 2	Method 1	Method 2
Mean bias of $\hat{\beta}^*$	0.00	-0.02	0.00	-0.20
95% QI	[-0.01; 0.01]	[-0.03; -0.01]	[-0.07; 0.08]	[-0.27; -0.14]
Mean empirical variance of X	4.77	4.67	5.37	4.66
Mean empirical variance of Z	4.68	4.76	4.68	6.41

Table 3 Mean bias, associated 95% quantile interval (95% QI) and mean empirical variance of the true exposure vector X and the observed exposure vector Z for Berkson error generated with a standard deviation of 0.8 and a true regression coefficient β of 1.

	Additive error		Multiplicative error	
	Method 1	Method 2	Method 1	Method 2
Mean bias of $\hat{\beta}^*$	0.00	-0.12	0.00	-0.77
95% QI	[-0.02; 0.02]	[-0.15; -0.09]	[-0.19; 0.27]	[-0.85; -0.72]
Mean empirical variance of X	5.3	4.68	11.33	4.67
Mean empirical variance of Z	4.66	5.32	4.68	40.41

As can be seen in these tables, Method 1 for the generation of Berkson error fulfills property 1 and property 2 as defined above since the mean bias is close to 0 and the variability of true exposure X_i exceeds the variability of observed exposure Z_i. Method 2 for the generation of Berkson error, on the other hand, does not fulfill these two properties. Indeed, the variability of observed exposure Z_i exceeds the variability of true exposure X_i and Berkson error generated according to this method introduces moderate to large biases in the estimation of the regression coefficient β.

4 Discussion

In the present simulation study, we showed that the use of Method 2 for the generation of Berkson error produced results that were not in accordance with two important properties of Berkson error. Contrary to Hoffmann et al. [9], Allodji et al. [7,8] found a substantial impact of measurement error in radon exposure when analyzing lung cancer mortality in the French cohort of uranium miners. This discrepancy could be due to the shortcomings of Method 2 for the generation of Berkson error.

These shortcomings can be explained by the fact that the independence assumption between observed exposure Z_i and the measurement error term U_i is violated when applying Method 2. Indeed, by construction, Z_i is a function of U_i (see equation (2)) and it therefore seems natural that Z_i and U_i cannot be independent. Due to the symmetry of the Gaussian distribution, in the additive case, it can be easily seen that Method 2 generates errors following unbiased classical measurement error. In the case of multiplicative error, on the other hand, it is necessary to take a closer look at the properties of the lognormal distribution. If a variable U is log-normally distributed with the mean and standard deviation of its natural logarithm μ and σ, respectively, then 1/U is also log-normally distributed, but the mean and the standard deviation of its natural logarithm are $-\mu$ and σ. Therefore, one can rewrite Method 2 to generate Berkson error as:

$$Z_i = X_i \cdot U_i' \tag{3}$$

where $U_i' = 1/U_i$. If $\log(U_i)$ follows a normal distribution with mean $-\frac{\sigma^2}{2}$ and standard deviation σ, then $\log(U_i')$ follows a normal distribution with mean $\frac{\sigma^2}{2}$ and standard deviation σ. Therefore, the expectation of U_i' is $\exp(\frac{\sigma^2}{2} + \frac{\sigma^2}{2}) = \exp(\sigma^2) \neq 1$ if $\sigma^2 \neq 0$, which violates the property of unbiased Berkson error that $E(U_i'|Z_i) = 1$. A comparison of expression (3) with the multiplicative classical measurement error model (see Table 1) shows that Method 2 for the generation of Berkson error inadvertently generates classical measurement error with a systematic bias (as $E(U_i) \neq 1$) instead of Berkson error without bias.

Table 4 Mean bias, associated 95% quantile interval (95% QI) and mean empirical variance of the true exposure vector X and the observed exposure vector Z for Berkson error generated by Method 2 and classical measurement error with a standard deviation of 0.3 and a true regression coefficient β of 1.

	Additive error		Multiplicative error	
	Method 2 Berkson	Unbiased classical	Method 2 Berkson	Biased classical
Mean bias of $\hat{\beta}^*$	-0.02	-0.02	-0.20	-0.20
95% QI	[-0.03; -0.01]	[-0.03; -0.01]	[-0.27; -0.14]	[-0.27; -0.14]
Mean empirical variance of X	4.67	4.67	4.66	4.67
Mean empirical variance of Z	4.76	4.76	6.41	6.42

Table 5 Mean bias, associated 95% quantile interval (95% QI) and mean empirical variance of true exposure X and observed exposure Z for Berkson error generated by Method 2 and classical measurement error with a standard deviation of 0.8 and a true regression coefficient β of 1.

	Additive error		**Multiplicative error**	
	Method 2 Berkson	Unbiased classical	Method 2 Berkson	Biased classical
Mean bias of $\hat{\beta}^*$	-0.12	-0.12	-0.77	-0.77
95% QI	[-0.15; -0.09]	[-0.15; -0.09]	[-0.85; -0.72]	[-0.85; -0.72]
Mean empirical variance of X	4.68	4.67	4.67	4.67
Mean empirical variance of Z	5.32	5.31	40.41	40.67

To verify these theoretical results, we can compare the results obtained for Method 2 with the situation where exposure data is generated with systematic multiplicative and unbiased additive classical measurement error. In the multiplicative case, we will generate $\log(U_i)$ according to a normal distribution with mean $\frac{\sigma^2}{2}$ and standard deviation σ. The results of this comparison for a standard deviation parameter of 0.3 and 0.8 are given in Table 4 and 5, respectively. These findings confirm the theoretical results that Berkson error generated according to Method 2 leads to unbiased classical measurement error in the additive case and to biased classical measurement error in the multiplicative case.

5 Conclusion

Berkson error plays an important role in epidemiology as many studies rely on a group-based exposure measurement strategy, for instance via job-exposure-matrices. Although it is often stated that Berkson error does not cause bias in risk estimates, this assumption is not true for logistic regression and proportional hazard models, in particular in the case of multiplicative errors [10] or in the case of errors which are correlated within or among study participants [11]. As it is indispensable to generate measurement error in order to develop and to validate methods for measurement error correction, our results underline the importance of a careful choice in the method of error generation.

References

1. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM, Measurement error in nonlinear models: a modern perspective, Chapman Hall, Boca Raton (2006)
2. Michels KB, A renaissance for measurement error, International Journal of Epidemiology, 30,421–422 (2001)
3. Simon SL, Hoffman FO, Hofer E, The two-dimensional Monte Carlo: a new methodological paradigm for dose reconstruction for epidemiological research,Radiation Research,183, 27–41 (2015)
4. Cook JR, Stefanski LA, Simulation-extrapolation estimation in parametric measurement error models, Journal of the American Statistical Association,89,1314–1328 (1994)
5. Stayner L, Vrijheid M, Cardis E, Stram DO, Deltour I, Gilbert SJ, et al, A Monte Carlo maximum likelihood method for estimating uncertainty arising from shared errors in exposures in epidemiological studies of nuclear workers, Radiation Research168, 757–763 (2007)
6. Kwon D, Hoffman FO, Moroz BE, Simon SL, Bayesian dose-response analysis for epidemiological studies with complex uncertainty in dose estimation, Statistics in medicine 35, 399–423 (2016)
7. Allodji RS, Leuraud K, Thiébaut AC, Henry S, Laurier D, Bénichou J, Impact of measurement error in radon exposure on the estimated excess relative risk of lung cancer death in a simulated study based on the French Uranium Miners' Cohort, Radiation and Environmental Biophysics 51, 151–163 (2012)
8. Allodji RS, Thiébaut A, Leuraud K, Rage E, Henry S, Laurier D, et al. The performance of functional methods for correcting non-Gaussian measurement error within Poisson regression: corrected excess risk of lung cancer mortality in relation to radon exposure among French uranium miners. Statistics in Medicine. 2012;31(30):4428–4443.
9. Hoffmann S, Rage E, Laurier D, Laroche P, Guihenneuc C, Ancelet S, Accounting for Berkson and classical measurement error in radon exposure using a Bayesian structural approach in the analysis of lung cancer mortality in the French cohort of uranium miners, Radiation Research, 187, 196–209 (2017)
10. Küchenhoff H, Bender R, Langner I, Effect of Berkson measurement error on parameter estimates in Cox regression models, Lifetime Data Analysis 13, 261–272 (2007)
11. Buonaccorsi JP, Lin CD, Berkson measurement error in designed repeated measures studies with random coefficients, Journal of Statistical Planning and Inference 104, 53–72 (2002)