Bundesamt
für Strahlenschutz

Ressortforschungsberichte zum Strahlenschutz

Ermittlung der Unsicherheiten in der Strahlenexpositionsabschätzung in der Wismut-Kohorte - Teil 2

Vorhaben 3618S12223

Ludwig-Maximilians-Universität München

N. Ellenbach
R. Rehms
Dr. S. Hoffmann

Das Vorhaben wurde mit Mitteln des Bundesministeriums für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (BMUV) und im Auftrag des Bundesamtes für Strahlenschutz (BfS) durchgeführt.

Impressum

Bundesamt für Strahlenschutz
Postfach 10 01 49
38201 Salzgitter

Tel.: +49 30 18333-0
Fax: +49 30 18333-1885
E-Mail: ePost@bfs.de
De-Mail: epost@bfs.de-mail.de

www.bfs.de

BfS-RESFOR-216/23

Bitte beziehen Sie sich beim Zitieren dieses Dokumentes immer auf folgende URN:
urn:nbn:de:0221-2023071238488

Salzgitter, Juli 2023

Abstract

Aim of the research project

Part 2 of the research project “Determination of uncertainties of radiation exposure assessment in the Wismut cohort” included the following tasks: (1) Quantification of uncertainty, (2) Definition of measurement models and development of an approach to correct for measurement error, (3) Design and implementation of a simulation study to compare the proposed approach with simulation extrapolation and regression calibration, and (4) Application to the data of the Wismut cohort without accounting for effect modifying variables and excluding workers who were employed in Wismut processing companies at any point during their working career.

Preliminary work from part 1

The project built on part 1 in which the working conditions in the Wismut company and procedures for estimating occupational exposure to radon progeny were described. Generalization error, assignment error, procedural measurement error, documentation error, parameter uncertainties, experts’ evaluation error, transfer error and approximation error were identified as potential sources of uncertainty in the Wismut cohort. In a preliminary evaluation, generalization error and parameter uncertainties were considered as particularly relevant.

Background

The Wismut cohort consists of a sample of 58 974 male employees from around 400 000 former employees of the Wismut company. The employees were exposed to various occupational exposures ranging from exposure to ionizing radiation through radon and its progeny, uranium dust and external gamma radiation to silica dust, arsenic and diesel exhaust. It constitutes one of the largest cohorts of uranium miners who were occupationally exposed to radon. When the cohort was established, individual exposure estimates for radon progeny were reconstructed through a Job Exposure Matrix (JEM) which provides information on the annual exposure for a hewer with 2000 working hours. In the early years of exposure in the Wismut cohort (1946 – 1954/55), there were no systematic exposure assessment, and exposure values received in this period therefore had to be reconstructed retrospectively by experts. Due to a lack of exposure information, it was however impossible to reconstruct the exposure values for each object and year independently. Starting in 1954/55, there was exposure monitoring for underground mining objects in the Wismut cohort based on measurements of radon gas concentration (1955/56 - 1965 in Saxony and 1955/56 - 1974 in Thuringia) and radon progeny concentration (1966 - 1990 in Saxony and 1975 - 1990 in Thuringia). In this exposure assessment period, measurements were taken in each year and object to estimate a mean annual radon gas concentration and radon progeny concentration, respectively. Radon gas or radon progeny estimates were multiplied by a working time factor, an activity weighting factor, and either an equilibrium factor (for radon gas concentration measurements) or a ventilation correction factor (for radon progeny concentration measurements).

Challenges of the research project

Estimating the association between time until death by lung cancer and cumulative radon progeny exposure in the Wismut cohort is challenging, because exposure is ongoing and time-dependent, rather than being a fixed point-exposure determined at study entry. Changes in the methods of exposure assessment create complex patterns of exposure uncertainty, where the type and magnitude of measurement error in the exposure history of a miner can vary over time. In a JEM, the same exposure level is assigned to all workers in a given year, object and activity. Measurement error in the estimation of this common exposure level will therefore affect all workers in that year, object

and activity in the same way. Moreover, individual exposure values for a worker are obtained by multiplying radon gas or radon progeny concentrations with several uncertain factors. Due to a lack of information, it was impossible to precisely estimate the values for these factors for each object and year, resulting in measurement errors that may simultaneously affect all years of a given mining object, or several mining objects, or both. It is important to account for these complex dependence structures as previous simulation studies show that shared measurement error components, particularly those in the retrospective assessment of exposure values for the earliest years of exposure, can lead to an attenuation of the exposure-response curve for high exposure values. This phenomenon, which is frequently observed in occupational cohort studies, can undermine the validity of risk estimates that ignore these complex patterns of measurement error.
Previous studies addressing the problem of measurement error in radon exposure in the Wismut cohort and in other cohorts of uranium miners have made a number of simplifying assumptions. The authors of these studies often 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, 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 exposure values received during the entire working career. Moreover, most of these studies used approaches to address exposure measurement error that make the simplifying assumption that measurement errors are unshared, i.e. that they are independent for different individuals and different exposure years of the same individual. By making this assumption, they neglect dependence structures in measurement errors arising in the estimation of exposure values through a JEM and are therefore not flexible enough to address the complex dependence structures in the measurement errors arising in the exposure assessment of the Wismut cohort.

Methods

Based on the preliminary evaluation made in part 1 of the research project, we derived a concept for the quantification of exposure uncertainty in the Wismut cohort. We mainly used information from the dosimetric reports from the Wismut company, estimates of exposure uncertainty in other cohorts of uranium miners and the information provided by Lehmann et al. (JEM1; 1998) and Lehmann (JEM 2; 2004) to derive estimates of exposure uncertainty for the Wismut cohort. To account for the complex measurement error components arising in the exposure assessment, we derived measurement error models for the different exposure assessment periods. Current approaches to account for measurement error lack the flexibility to account for these complex measurement error models. We chose a Bayesian hierarchical approach, which is based on the combination of sub models via conditional independence assumptions, as it provides a flexible and coherent framework for the treatment of complex phenomena which may be prone to multiple sources of uncertainty. Bayesian inference for these models was conducted using an object-oriented implementation of a Markov chain Monte Carlo algorithm in Python. In a simulation study, the performance of the proposed approach was compared to the more classical approaches simulation extrapolation and regression calibration.

Results

The proposed Bayesian hierarchical approach showed very good performance on simulated data with a relative bias of -2.98%, 6.76% and 3.50% in the different simulation scenarios and coverage rates that were very close to the nominal level of 95%. For the full Wismut cohort, we estimated an excess hazard ratio of 0.54 per 100 WLM with a 95% credible interval of [0.35, 0.81] compared to a naive excess hazard ratio of 0.33 per 100 WLM [0.27, 0.40] when measurement error is ignored. For the

cohort of workers hired in 1960 or later, the estimated excess hazard ratio was 1.80 per 100 WLM [0.70, 3.36] compared to a naive estimate of 1.44 per 100 WLM [0.73, 2.52]. These results are preliminary, as they do not yet account for all measurement models that are to be considered and their robustness still has to be confirmed in extensive sensitivity analyses. Moreover, they do not include data on workers who were employed in processing companies at any point during their working career.

Discussion

While the Bayesian hierarchical approach shows great flexibility accounting for complex patterns of measurement error, the results of the current work have to be interpreted with caution. Indeed, they rely on many assumptions on the magnitude and the structure of measurement error. It was not in the scope of this project to conduct extensive sensitivity analyses to assess the robustness of the results to assumptions on the structure and magnitude of measurement error in the Wismut cohort. Moreover, it seems essential to consider effect modifying variables in the association between radon exposure and lung cancer mortality in the full Wismut cohort. On the one hand, a previous simulation study (Hoffmann et al., 2018b) suggests that complex structures of measurement error may lead to apparent effect modification when the true model is linear without effect modification. On the other hand, if the true model includes effect modifying variables and a simple linear model without effect modification is assumed, this model misspecification may interfere with the correction of measurement error. Finally, it would be important to refine the model for the first exposure assessment periods in the Wismut cohort in future work as the measurement errors arising in this period were both, very complex and very large.

Contents

1 Introduction

Exposure measurement error poses an important threat to the validity of statistical inference in occupational epidemiology. 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 (Carroll, 2005; Blair et al., 2011; Carroll et al., 2006).

This project is concerned with the quantification and correction for exposure measurement error in the Wismut cohort. This cohort consists of a sample of 58 974 male employees from around 400 000 former employees of the Wismut company. It constitutes one of the largest cohorts of uranium miners who were occupationally exposed to radon progeny. The employees were exposed to various occupational exposures ranging from exposure to ionizing radiation through radon and its progeny, uranium dust and external gamma radiation to silica dust, arsenic and diesel exhaust. When the cohort was established, individual exposure estimates for radon progeny were assessed through a Job Exposure Matrix (JEM) which provides information on the annual exposure for a hewer with 2000 working hours. In the early years of exposure in the Wismut cohort (1946 – 1954/55), there were no systematic exposure assessment and exposure values received in this period therefore had to be reconstructed retrospectively by experts. Due to a lack of systematic exposure measurements, it was however impossible to reconstruct the exposure values for each object and year independently. Starting in 1954/55, there was exposure monitoring for underground mining objects in the Wismut cohort based on measurements of radon gas concentration (1955/56 - 1965 in Saxony and 1955/56 - 1974 in Thuringia) and radon progeny concentration (1966 - 1990 in Saxony and 1975 - 1990 in Thuringia). In this exposure assessment period, measurements were taken in each year and object to estimate a mean annual radon gas concentration and radon progeny concentration, respectively. Radon gas or radon progeny concentrations were calculated using a working time factor, an activity weighting factor, and either an equilibrium factor (for radon gas concentration measurements) or a ventilation correction factor (for radon progeny concentration measurements).

Estimating the association between time until death by lung cancer and cumulative radon exposure in the Wismut cohort is challenging, because cumulative radon exposure is ongoing and time-dependent, rather than being a fixed point exposure determined at study entry. Changes in the methods of exposure assessment create complex patterns of exposure uncertainty, where the type and magnitude of measurement error in the exposure history of a miner can vary over time. In a JEM, the same exposure level is assigned to all workers in a given year, object and activity. Measurement error in the estimation of this common exposure level will therefore affect all workers in that year, object and activity in the same way. Moreover, individual exposure values for a worker are obtained using radon gas or radon progeny concentrations and several uncertain factors. Due to a lack of information, it was impossible to estimate the values for these factors for each object and year, resulting in measurement errors that may simultaneously affect all years of a given mining object, or several mining objects, or both. It is important to account for these complex dependence structures as previous simulation studies show that shared error components, particularly those in the retrospective assessment of exposure values for the earliest years of exposure, can lead to an attenuation of the exposure-response curve for high exposure values. This phenomenon, which is frequently observed in occupational cohort studies, can undermine the validity of risk estimates that ignore these complex patterns of measurement error.

Current approaches to address exposure measurement error classically assume that errors are unshared, i.e. that they are independent for different individuals and different exposure years of the same individual. They are therefore not flexible enough to address the complex dependence structures in the errors arising in the exposure assessment of the Wismut cohort. Previous studies addressing the problem of measurement error in radon exposure in the Wismut cohort and in other cohorts of uranium miners have made a number of simplifying assumptions. The authors of these studies often 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, 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 in an exposure assessment via JEMs can be described by unshared measurement error, thereby neglecting the dependence structures in measurement errors arising in the estimation of exposure values through a JEM.
The current project builds on part 1 “Determination of uncertainties of radiation exposure assessment in the Wismut cohort” (Küchenhoff et al., 2018) in which the working conditions at the Wismut company and procedures for estimating occupational exposure to radon progeny were described. In Küchenhoff et al. (2018), generalization error, assignment error, procedural measurement error, documentation error, parameter uncertainties, experts’ evaluation error, transfer error and approximation error were identified as potential sources of uncertainty in the Wismut cohort. The aims of the current project are (1) Quantification of uncertainty, (2) Definition of measurement error models and development of an approach to correct for measurement error, (3) Design and implementation of a simulation study to compare the proposed approach with simulation extrapolation and regression calibration and (4) Application to the data of the Wismut cohort without accounting for effect modifying variables and excluding workers who were employed in processing companies at any point during their working career.
Chapter *2* gives a short overview of the methods of exposure assessment in the Wismut cohort and then describes these exposure characteristics through a complex measurement model in which the type and magnitude of error vary depending on the exposure assessment period and the type of object. Chapter *3* first provides a detailed concept for the quantification of exposure uncertainty in the Wismut cohort and then shows the results of this quantification to the extent that it has been carried out in the scope of this project. Chapter *4* presents the Bayesian hierarchical approach that was chosen in this project to account for the complex patterns of exposure measurement error arising in the Wismut cohort. Chapter *5* describes the design and the result of a simulation study that was conducted to assess the performance of the proposed Bayesian hierarchical approach and to compare it with simulation extrapolation and regression calibration. Chapter *6* provides the results when accounting for measurement error on the data of the Wismut cohort and Chapter *7* discusses these results and proposes aspects that might be relevant for future projects aiming to account for exposure uncertainty in the Wismut cohort.

2 Measurement models in the Wismut cohort

2.1 Exposure assessment in the Wismut cohort: the general approach

The methods of exposure assessment in the Wismut cohort changed over the years. For all exposure years, individual exposure estimates were based on a JEM which provides information on the annual exposure values to radon progeny for a hewer with 2000 working hours (Küchenhoff et al., 2018). This JEM initially contained object specific exposure values (JEM 1; Lehmann et al., 1998) which were later adapted to provide shaft specific exposure estimates (JEM 2; Lehmann, 2004). Following Küchenhoff et al. (2018), we will use the notation:

- object o (the same naming and numbering of the objects as in Küchenhoff et al. (2018) is used: see Table 2 in Küchenhoff et al. (2018, pp. 15–18))
- reference object o_0
- year t
- reference year t_0
- period p_t or $p(t, o)$ (the respective periods are different for different parameters)
- activity j
- reference activity j_0
- true exposure X
- observed exposure Z

In general, in the following sections, we first describe how the exposure to radon progeny for year t, object o and a worker conducting activity j $E(t, o, j)$ is calculated for the different object types and time periods.

To obtain exposure estimates for individual miners, this $E(t, o, j)$ is further multiplied by $l(i, t, o, j)$, reflecting the individual working history of miner i to obtain the individual observed exposure $Z_i(t, o, j) = E(t, o, j) \cdot l(i, t, o, j)$. If a miner worked in different objects and activities in a given year, the total true or observed exposure that he received in this given year is obtained by taking the sum over all objects and activities he worked in, $X_i(t) = \sum_{o,j} X_i\,(t, o, j)$ or $Z_i(t) = \sum_{o,j} Z_i\,(t, o, j)$, respectively.

Based on the exposure assessment described in Chapter 4 of Küchenhoff et al. (2018), Figure *2.1* shows the object structure in the Wismut and the corresponding measurement models, which we will explain in the following sections. Following Küchenhoff et al. (2018, p. 19), the frequency of occurrence is shown using the proportion of person work years (PPY), which is defined as

$$\text{PPY} = \frac{\text{Number of person work years in specific subgroup}}{\text{Total number of person work years in the Wismut cohort}}.$$

Note that the data set used at the beginning of this project for Figure *2.1* was a preliminary grouped data set due to data protection requirements. The main purpose of Figure *2.1* is to depict the structure of the objects and the corresponding measurement models and the values of the PPY are not final.

We will distinguish the following measurement models according to different object types and different exposure assessment periods. For underground mining objects the exposure assessment can be divided into three periods (experts' estimation, radon gas concentration measurements, radon progeny concentration measurements) and for processing companies into other two periods (experts' estimation, radon gas concentration measurements). Unless otherwise stated, in the following we will always refer to the exposure assessment periods of underground mining objects.

- M1a: Underground mining objects as well as exploration and development objects in Saxony in the first exposure assessment period (1946-1954/55)

- M1b: Underground mining objects as well as development objects in Thuringia in the first exposure assessment period (1946-1954/55)
- M2: Underground mining objects in Saxony and Thuringia as well as development objects in Saxony in the second exposure assessment period (1955/56-1965 in Saxony, 1955/56-1974 in Thuringia)
- M3: Underground mining objects in Saxony and Thuringia as well as development objects in Saxony in the third exposure assessment period (1966-1990 in Saxony, 1975-1990 in Thuringia)
- M4: Surface areas affiliated to mining objects and exploration objects in Thuringia (1946-1990)
- M5a: Processing companies in the first exposure assessment period for processing (1950-1962 for processing facilities, 1950-1960 for collieries)
- M5b: Processing companies in the second exposure assessment period for processing (1963-1990 for processing facilities, 1961-1980 for collieries, 1959-1990 for RAS and RAF facilities)
- M6: Open pit mining objects (1946-1990)

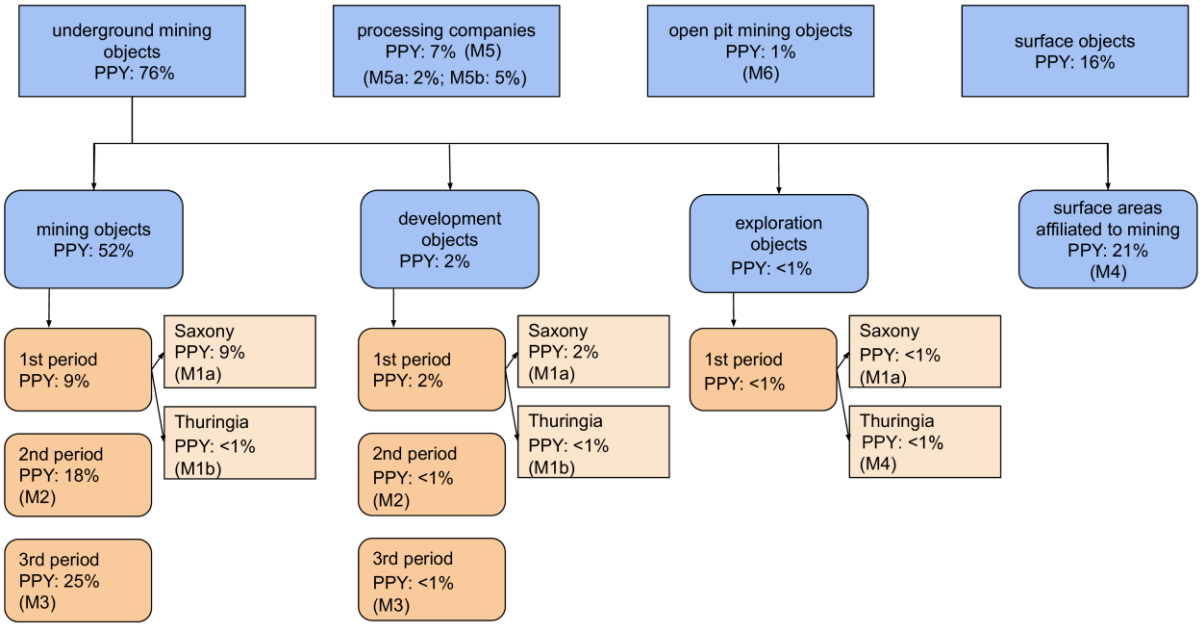

Figure 2.1: Overview of the object structure in the Wismut cohort with the respective measurement models to describe exposure uncertainty and the PPY based on a preliminary grouped data set.

For each of the different object types in the Wismut we will first describe the exposure assessment and then assign the corresponding measurement model.

In principle we could account for uncertainty in all intervening quantities. However, for some quantities, for instance the mined vein area $C(t, o)$ and the uranium recovering rate $r(t, o)$, it would be infeasible to quantify the uncertainty. Also, it can be assumed that these quantities can be measured with sufficient precision to consider them as known without measurement error. Table *2.1* depicts all quantities of the exposure assessment in the different object types of the Wismut, that we assume to be known without measurement error.

To visualize the exposure assessment and the respective measurement model we will use directed acyclic graphs (DAGs; Jordan, 2004). With circles we represent unknown quantities and with boxes observed quantities. Arrows indicate the dependencies between the different quantities, where a single arrow shows a probabilistic, and double arrows show a deterministic dependency. In order to make it clear how the respective measurement models can be accounted for, we will always show a hierarchical model in which this measurement model is combined with a disease model. The latter describes how the right censored variable time until death by lung cancer, represented through

$\{Y_i, \delta_i\}$, is modelled as a function of true cumulative exposure of miner i until time t $X_i^{cum}(t)$ in a proportional hazards model. In this disease model, we assume a piecewise constant baseline hazard with parameters $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and the association between cumulative radon exposure and lung cancer mortality is quantified through the parameter β. For more details on the disease model, see Section 4.4.1.

Table 2.1: Quantities of the different measurement models assumed to be known without measurement error.

Measurement model	Quantity assumed to be known without measurement error
M1a	$C(t, o)$: mined vein area
M1a	$r(t, o)$: relative uranium recovery rate
M1a, M1b	$A(t_0(o_0(o)), o_0(o))$: evaluation area of the reference object and respective reference year
M1b	$A(t, o)$: evaluation area
M6	$d(t, o)$: depth
M1a, M1b, M2, M3, M4, M5a, M5b, M6	$l(i, t, o, j)$: individual working history of miner i

2.2 Uncertainties in exposure assessment based on experts' estimation in underground mining objects (1946 - 1954/55)

In the exposure assessment of underground mining objects in the Wismut cohort (more precisely, for mining objects (PPY: 52%), development objects (PPY: 2%) and exploration objects (PPY: <1%)), it is possible to distinguish three different periods, which will be described in this and the following two sections: experts' estimation (PPY: 11%), estimation based on radon gas concentration measurements (PPY: 18%; Section *2.3*) and estimation based on radon progeny concentration measurements (PPY: 25%; Section *2.4*).

As systematic radon measurements for radiation protection in the Wismut only started in 1954 (Küchenhoff et al., 2018, p. 34), exposure values before this year could not be based on direct measurements and therefore had to be reconstructed retrospectively by experts. This exposure assessment is described in detail by Lehmann et al. (1998) and Küchenhoff et al. (2018).
In the following, we will summarize the most important sources of uncertainty in the exposure assessment in this period that were identified by Küchenhoff et al. (2018).

2.2.1 General radon gas exposure assessment

In the exposure assessment, the annual exposure to radon gas of a hewer in underground mining with 2000 working hours per year $E^*(t, o, j_0(o))$ for the mining objects is assumed to be the sum of basic exposure from old mining $E^B(o)$ for object o and from mining activity $E^M(t, o, j_0(o))$:

$$E^*(t, o, j_0(o)) = E^B(o) + E^M(t, o, j_0(o)).$$

Basic exposure from old mining

For all old mining objects, the value for the basic exposure from old mining $E^B(o)$ was calculated as a proportion $b(o)$ of the basic exposure to radon gas which had been estimated for object 003 Schneeberg through measurements performed in 1937/1938 by Rajewski (Küchenhoff et al., 2018).

Based on a measurement of 22.5 Eman by Rajewski (Küchenhoff et al., 2018), an assumed equilibrium factor $g(t,o)$ of 0.6 (Lehmann et al., 1998) and a working time factor of 1.2, one obtains a value of $E^B(003) = 12 \cdot 22.5 \cdot 0.6 \cdot 1.2 = 194$ WLM for object 003 Schneeberg. This value was based on 70 measurements (Küchenhoff et al., 2018). For all new ground-opening objects, the basic exposure from old mining was supposed to be 0.

Exposure from mining activity

The annual exposure to radon gas from mining activity $E^M(t,o,j_0(o))$ in year t and object o for the reference activity $j_0(o)$ in this object was calculated by multiplying the evaluation area $A(t,o)$ and the evaluation factor $e(o)$ for objects in Thuringia. For objects in Saxony, this product was further multiplied by the relative uranium recovering rate $r(t,o)$.
For Saxony:

$$E^M(t,o,j_0(o)) = A(t,o) \cdot e(o) \cdot r(t,o)$$

For Thuringia, there was no exposure from old mining, i.e. $E^*(t,o,j_0(o)) = E^M(t,o,j_0(o))$, leading to the following equation:

$$E^*(t,o,j_0(o)) = E^M(t,o,j_0(o)) = A(t,o) \cdot e(o)$$

The three main quantities intervening in the assessment of exposure values are therefore:

- the evaluation area $A(t,o)$ as a measure for the size of the radon exit field
- the evaluation factor $e(o)$ as a measure of the exposure to radon per unit of the mined area
- the relative uranium recovery rate $r(t,o)$ (only for objects in Saxony) as a measure of the uranium content of the bedrock of object o in relation to the reference object 009 Aue in the reference year 1955

Evaluation area $A(t,o)$ in Saxony

As the uranium mineralization in Saxony was bounded by vein structures, the evaluation area was approximated through the mined vein area $C(t,o)$ (Küchenhoff et al., 2018, p. 48). In the estimation of the evaluation area in Saxony, it was assumed that the vein area which had been mined in the previous years had a non-negligible influence on the evaluation area but there was some uncertainty on the extent of this influence (Lehmann et al., 1998, p. 66). In accordance with the loss in the first years, which was approximately 20%, it was assumed that the area mined in the previous years should be weighted by a factor $p = 0.2$ in the estimation of the evaluation area $A(t,o)$ in Saxony.

$$A(t,o) = C(t,o) + p \sum_{s=1946}^{t-1} C\,(s,o)$$

Evaluation area $A(t,o)$ in Thuringia

As the uranium mineralization in Thuringia was not bounded by vein structures, the evaluation area was approximated through the void volume $V(t,o)$ (Küchenhoff et al., 2018, p. 48). The total void volume created before 1955 was unknown and had to be estimated through the total shaft output of object o before 1955 $F(o) = \sum_{u=1946}^{1955} F\,(u,o)$ divided by the density of the bedrock $h(o)$ (Lehmann et al., 1998, p. 118). In order to derive the void volume created in every year before 1955, the relative uranium recovery $q(t,o)$ of object o in year t has to be determined. To do this, the uranium recovery in year t and object o $R(t,o)$ has to be divided by the total uranium recovery of object o between 1946 and 1955 $R(o) = \sum_{u=1946}^{1955} R\,(u,o)$

$$q(t,o) = \frac{R(t,o)}{\sum_{u=1946}^{1955} R\,(u,o)} = \frac{R(t,o)}{R(o)}.$$

Finally, the void volume created in object o in year t is obtained as

$$V(t,o) = \frac{\sum_{u=1946}^{1955} F\,(u,o)}{h(o)} q(t,o) = \frac{F(o)}{h(o)} q(t,o)$$

yielding the following formula to estimate the estimation area $A(t,o)$ of object o in year t for objects in Thuringia:

$$A(t,o) = \sum_{s=1946}^{t} V\,(s,o) = \frac{\sum_{u=1946}^{1955} F\,(u,o)}{h(o)} \frac{\sum_{s=1946}^{t} R\,(s,o)}{\sum_{u=1946}^{1955} R\,(u,o)} = \frac{F(o)}{h(o) \cdot R(o)} \sum_{s=1946}^{t} R\,(s,o).$$

Evaluation factor $e(o)$

In order to determine an estimate of $E^M(t,o,j_0(o))$, the evaluation area $A(t,o)$ has to be multiplied by a factor which quantifies the annual exposure to radon gas per unit of the mined area. To derive this factor, which is referred to as evaluation factor $e(o)$, measurements of radon gas concentration were used which had been obtained in 1955 in 006 Vogtland-Zobes, 009 Aue and 903 BB Schmirchau and in 1962 in 904 BB Paitzdorf, which were considered as reference objects for the estimation of $e(o)$. To derive the evaluation factors $e(o)$ of these four objects, the arithmetic mean of the measurements in each object was multiplied by 12 and divided by the evaluation area $A(t,o)$ of the corresponding year and object. For all other objects in this exposure assessment period, values of $e(o)$ were determined based on expert knowledge by considering the evaluation factors determined for the four reference objects and mine ventilation conditions and the uranium mineralization in comparison with the four reference objects. Thus, the evaluation factor for non-reference objects can be obtained by multiplying the evaluation factor of the respective reference object with a transfer factor for the evaluation factor $t_e(o)$.

$$\begin{aligned} e(o) = &\quad e(o_0(o)) \cdot t_e(o) \\ = &\quad \frac{C_{Rn}(t_0(o_0(o)), o_0(o)) \cdot 12}{A(t_0(o_0(o)), o_0(o))} \cdot t_e(o) \end{aligned}$$

Relative uranium recovering rate $r(t,o)$ for objects in Saxony

To obtain an assessment of the exposure value $E^M(t,o,j_0(o))$ for objects in Saxony, the product between the evaluation area $A(t,o)$ and the evaluation factor $e(o)$ needs to be multiplied with a further quantity. This quantity is referred to as the relative uranium recovering rate $r(t,o)$ and measures the uranium content of the bedrock in comparison to the reference object 009 Aue in the reference year 1955 (Küchenhoff et al., 2018, p. 50).

2.2.2 Exposure assessment for objects in Saxony

As described previously and on pages 46 to 51 in Küchenhoff et al. (2018), in the years 1946 to 1954/55 exposure to radon progeny for year t, object o and a worker conducting activity j in Saxony was estimated according to the following formula (parameter definitions are given on the next pages):

$$\begin{aligned} E(t,o,j) = &\quad (C_{Rn}(1937/1938{,}003) \cdot b(o) \cdot 12 + \\ &\quad r(t,o) \cdot e(o) \cdot t_e(o) \cdot \\ &\quad A(t,o)) \cdot \\ &\quad g(p_t,o) \cdot w(p_t) \cdot f(o,j) \\ = &\quad (C_{Rn}(1937/1938{,}003) \cdot b(o) + \\ &\quad r(t,o) \cdot \frac{C_{Rn}(t_0(o_0(o)), o_0(o))}{A(t_0(o_0(o)), o_0(o))} \cdot t_e(o) \cdot \\ &\quad \left(C(t,o) + p \sum_{s=1946}^{t-1} C\,(s,o) \right) \cdot \\ &\quad g(p_t,o) \cdot w(p_t) \cdot f(o,j) \cdot 12 \end{aligned}$$

As described in Küchenhoff et al. (2018, p. 54), the radon exposure in exploration objects was lower than in underground mining objects since there was less ore contact. The exposure assessment depends on the region. For Saxonian exploration objects, the radon exposure was assessed by multiplying the radon progeny exposure of the respective reference underground mining object with the factor t_E.

$$E(t, o, j) = t_E \cdot E(t, o_0(o), j)$$

For exploration objects in Saxony with new-ground opening reference underground mining objects the factor is $t_E = 0.1$ and for those Saxonian exploration objects for which old underground mining objects were the reference object, the factor is $t_E = 0.2$. For development objects in Saxony, the radon exposure was assessed the same way as the exposure in exploration objects in Saxony but with a different value of the factor t_E. Development objects in Saxony use the factor $t_E = 0.3$.

The exposure estimation for objects in Saxony in the first exposure assessment period is illustrated in the DAG shown in Figure *2.2*. The transfer factor t_E that is depicted in the DAG in Figure *2.2* is only relevant for exploration and development objects in Saxony. For underground mining objects that are no exploration or development objects, the transfer factor is set to $t_E = 1$.

For Saxonian underground mining, exploration and development objects in the first exposure assessment period, the parameters are defined as follows:

- $f(o, j)$: activity weighting factor
- $b(o)$: proportion of exposure from old mining in comparison to object 003 Schneeberg
- $w(p_t)$: working time factor to adjust for the actual working time of a hewer
- $g(p_t, o)$: equilibrium factor as a measure of the disequilibrium of radon and its progeny, which is used to convert a measure of radon gas exposure to radon progeny exposure in working level months (WLM), which is the historical unit of radon exposure in cohorts of uranium miners and related to the potential alpha energy concentration (Marsh et al., 2012).
- Measurements of radon gas
  - $C_{Rn}(1937/1938{,}003)$: mean concentration measurement of radon gas for object 003 Schneeberg in 1937/1938 intervening in the estimation of radon exposure due to old mining for all old mining objects
  - $C_{Rn}(1955{,}006)$ and $C_{Rn}(1955{,}009)$: mean concentration measurements of radon gas for objects 006 Vogtland-Zobes and 009 Aue in 1955 were used to estimate the evaluation factors $e(o)$ for these two objects
- p: 20% of the cumulative mined area $C(s, o)$ in the previous years $s \in \{1946, \dots, t-1\}$ intervene in the calculation of the evaluation area $A(t, o)$ of year t for objects in Saxony (Küchenhoff et al., 2018, p. 48)
- $t_e(o)$: multiplicative factor to determine the evaluation factor of non-reference objects as a function of the reference objects (Küchenhoff et al., 2018, p. 50; Lehmann et al., 1998, p. 67)
- t_E: multiplicative transfer factor to calculate the exposure to radon progeny for exploration and development objects based on the exposure of the reference mining object (only for exploration and development objects)
- $A(1955{,}006)$ and $A(1955{,}009)$: evaluation area for objects 006 Vogtland-Zobes and 009 Aue in 1955 were used to estimate the evaluation factors $e(o)$ for these two objects
- $r(t, o)$: relative uranium recovery rate
- $R(t, o)$: amount of uranium recovery
- $C(t, o)$: mined vein area
- $A(t, o)$: evaluation area

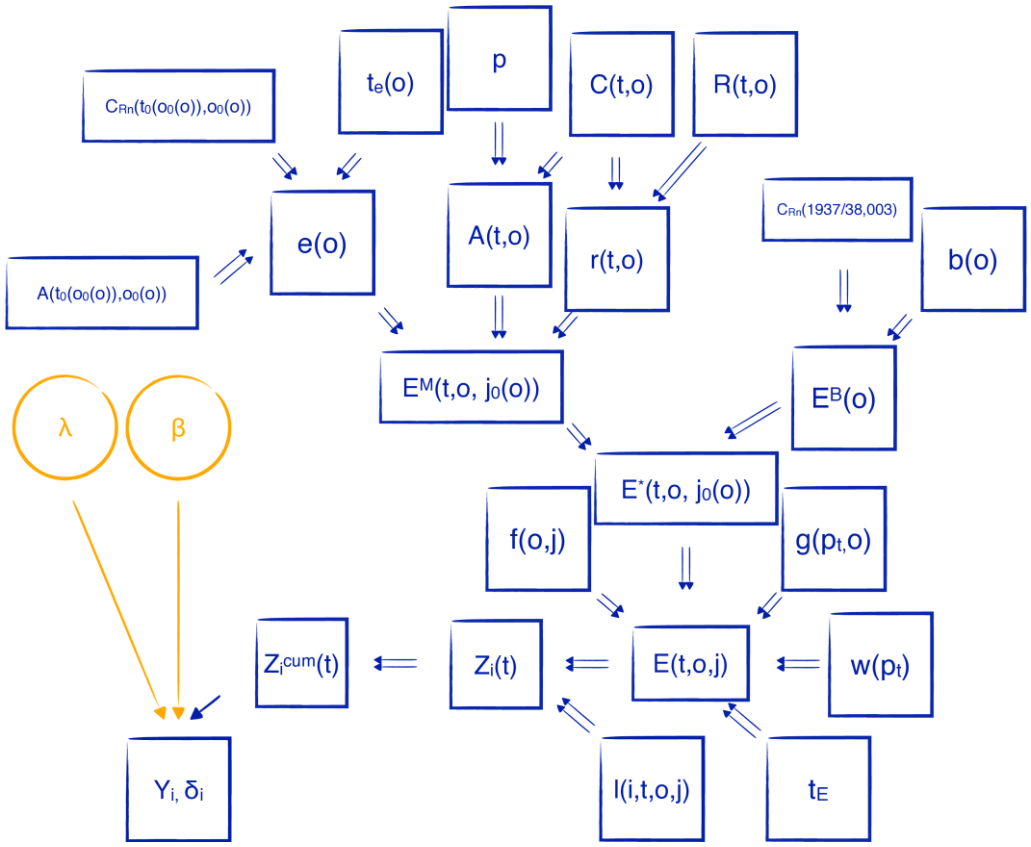

Figure 2.2: Exposure assessment in underground mining objects as well as in exploration and development objects in Saxony in the first exposure assessment period.

2.2.3 Measurement model M1a for objects in Saxony

For the uncertainties arising in the exposure estimation for underground mining, development and exploration objects in Saxony in the first exposure assessment period we assume measurement model M1a shown in Figure *2.3* in which the measurement error structure becomes evident. Basically, we use normal Latin letters for the observed parameters, while we use italic or Greek letters for the erroneous, latent parameters. For differentiation, the parameters affected by Berkson error are shown with an apostrophe, whereas the true mean parameters are shown without.

We assume a classical measurement error for the mean radon gas concentration measurements $C_{Rn}(t_0(o_0(o)), o_0(o))$ and $C_{Rn}(1937/1938{,}003)$. For the remaining uncertain quantities intervening in the exposure assessment for objects in Saxony in the first exposure assessment period, we assume a classical and a Berkson error component.
As mentioned in more detail in Section *3.1*, many authors argue in favor of a multiplicative error component that follows a lognormal distribution for measurement error in radiation exposure in general, and in radon exposure in particular (Lubin et al., 1995b; Stram et al., 1999; Heid, 2002; Heid et al., 2002; Heid et al., 2004; Heidenreich et al., 2004; Lubin et al., 2005; Advisory Group on Ionising Radiation AGIR, 2009; Heidenreich et al., 2012; Allodji et al., 2012a,b,c). The multiplicative error structure moreover has the advantage of naturally respecting the non-negativity of true exposure values for observed exposure values that are non-negative. In the following, we will therefore assume multiplicative errors for all parameter uncertainties. Concerning the errors arising in mean radon gas and radon progeny concentrations, however, we will follow Küchenhoff et al. (2018) who assume an additive error structure whenever the measurements were the result of an averaging of measurements.

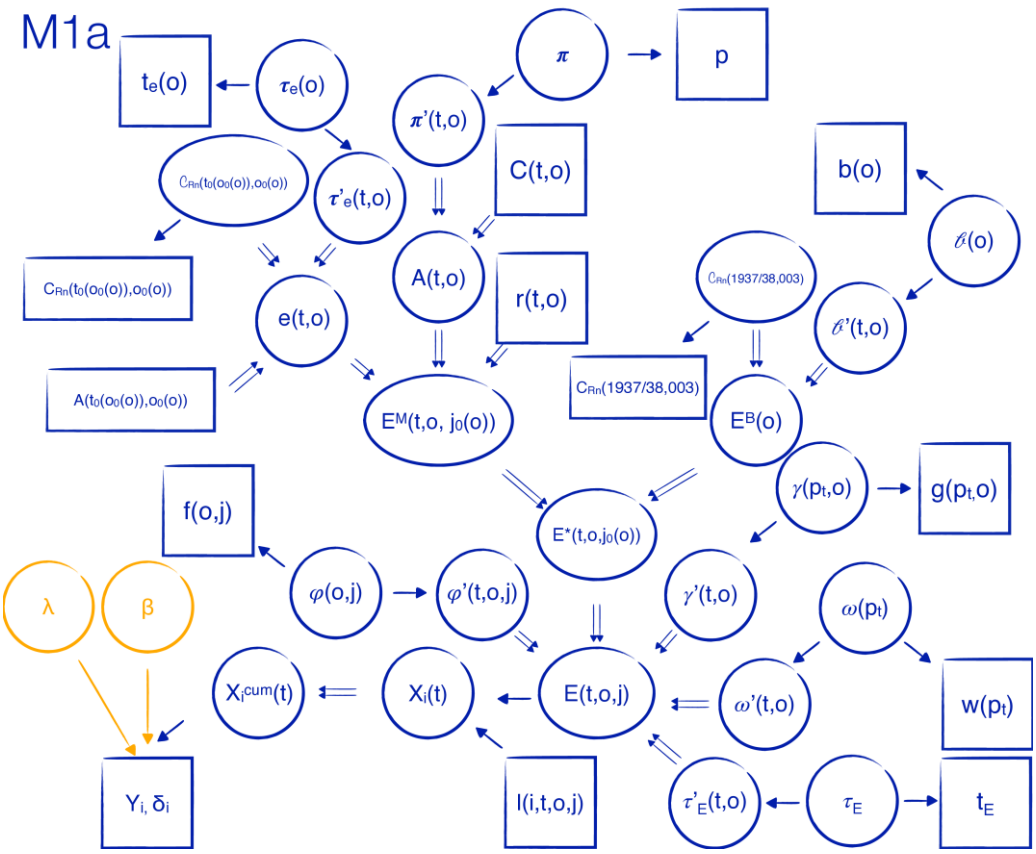

Figure 2.3: Hierarchical model combining a disease model with measurement model M1a to describe exposure uncertainty in underground mining objects as well as in exploration and development objects in Saxony in the first exposure assessment period. Due to the limited space and for a clearer presentation, no measurement error variances are shown here.

Thus, for underground mining, development and exploration objects in Saxony in the first exposure assessment period, we assume the following measurement model, where the uncertainty in the transfer factor t_E is only accounted for in exploration and development objects in Saxony.

$$
\begin{aligned}
C_{Rn}(t_0(o_0(o)), o_0(o)) &= \mathcal{C}_{Rn}(t_0(o_0(o)), o_0(o)) + U_{\mathcal{C},c}(t_0(o_0(o)), o_0(o)) \\
C_{Rn}(1937/1938{,}003) &= \mathcal{C}_{Rn}(1937/1938{,}003) + U_{\mathcal{C},c}(1937/1938{,}003) \\
f(o,j) &= \varphi(o,j) \cdot U_{\varphi,c}(o,j) \\
\varphi'(t,o,j) &= \varphi(o,j) \cdot U_{\varphi',B}(t,o,j) \\
w(p_t) &= \omega(p_t) \cdot U_{\omega,c}(p_t) \\
\omega'(t,o) &= \omega(p_t) \cdot U_{\omega',B}(t,o) \\
g(p_t,o) &= \gamma(p_t,o) \cdot U_{\gamma,c}(p_t,o) \\
\gamma'(t,o) &= \gamma(p_t,o) \cdot U_{\gamma',B}(t,o) \\
b(o) &= \mathcal{b}(o) \cdot U_{\mathcal{b},c}(o) \\
\mathcal{b}'(t,o) &= \mathcal{b}(o) \cdot U_{\mathcal{b}',B}(t,o) \\
p &= \pi \cdot U_{\pi,c} \\
\pi'(t,o) &= \pi \cdot U_{\pi',B}(t,o) \\
t_e(o) &= \tau_e(o) \cdot U_{\tau_e,c}(o) \\
\tau_e{}'(t,o) &= \tau_e(o) \cdot U_{\tau_e',B}(t,o) \\
t_E &= \tau_E \cdot U_{\tau_E,c} \\
\tau_E{}'(t,o) &= \tau_E \cdot U_{\tau_E',B}(t,o)
\end{aligned}
$$

In summary, while the observed exposure $Z_i(t,o,j)$ for miner i in year t, object o and with activity j in the first exposure assessment period for underground mining objects in Saxony was calculated as

$$\begin{aligned} Z_i(t,o,j) = \quad & (C_{Rn}(1937/1938{,}003)\cdot b(o) + \\ & r(t,o)\cdot\frac{C_{Rn}(t_0(o_0(o)),o_0(o))}{A(t_0(o_0(o)),o_0(o))}\cdot t_e(o)\cdot \\ & \left(C(t,o)+p\sum_{s=1946}^{t-1}C\,(s,o)\right)\Bigg)\cdot t_E\cdot \\ & g(p_t,o)\cdot w(p_t)\cdot f(o,j)\cdot 12\cdot l(i,t,o,j), \end{aligned}$$

the true value $X_i(t,o,j)$ for this exposure depends on the unknown quantities defined in the previous equations in the following way:

$$\begin{aligned} X_i(t,o,j) = \quad & (\mathcal{C}_{Rn}(1937/1938{,}003)\cdot \mathcal{b}'(t,o) + \\ & r(t,o)\cdot\frac{\mathcal{C}_{Rn}(t_0(o_0(o)),o_0(o))}{A(t_0(o_0(o)),o_0(o))}\cdot {\tau_e}'(t,o)\cdot \\ & \left(C(t,o)+\pi'(t,o)\sum_{s=1946}^{t-1}C\,(s,o)\right)\Bigg)\cdot {\tau_E}'(t,o)\cdot \\ & \gamma'(t,o)\cdot\omega'(t,o)\cdot\varphi'(t,o,j)\cdot 12\cdot l(i,t,o,j)+U_{E,B}(i,t) \end{aligned}$$

where $l(i,t,o,j)$ reflects the individual working history of miner i.

2.2.4 Exposure assessment for objects in Thuringia

As described previously in this section and on pages 46 to 51 in Küchenhoff et al. (2018), in the years 1946 to 1954/55 exposure to radon progeny for year t, object o and a worker conducting activity j in Thuringia was estimated according to the following formula:

$$\begin{aligned} E(t,o,j) = \quad & (e(o)\cdot t_e(o) \\ & \cdot A(t,o))\cdot \\ & g(p_t,o)\cdot w(p_t)\cdot f(o,j) \\ = \quad & \left(\frac{C_{Rn}(t_0(o_0(o)),o_0(o))}{A(t_0(o_0(o)),o_0(o))}\cdot t_e(o)\right. \\ & \left.\cdot\left(\frac{F(o)}{h(o)\cdot R(o)}\sum_{s=1946}^{t}R\,(s,o)\right)\right)\cdot \\ & g(p_t,o)\cdot w(p_t)\cdot f(o,j)\cdot 12 \end{aligned}$$

For exploration objects in Thuringia the radon exposure was set to the value of 1 WLM, since no exploitable uranium mineralization was found. This reconstruction of radon exposure is very similar to the exposure assessment of surface areas affiliated to mining objects and is thus included into Section *2.5*. However, for development objects in Thuringia the radon exposure was assessed the same way as the exposure in development objects in Saxony with a value of $t_E = 0.3$.

The exposure estimation for underground mining and development objects in Thuringia in the first exposure assessment period is illustrated in the DAG shown in Figure *2.4*. Similar to the exposure assessment in underground mining objects in Saxony, the transfer factor t_E in Figure *2.4* is only relevant for development objects in Thuringia and is set to $t_E = 1$ for the remaining underground mining objects in Thuringia.

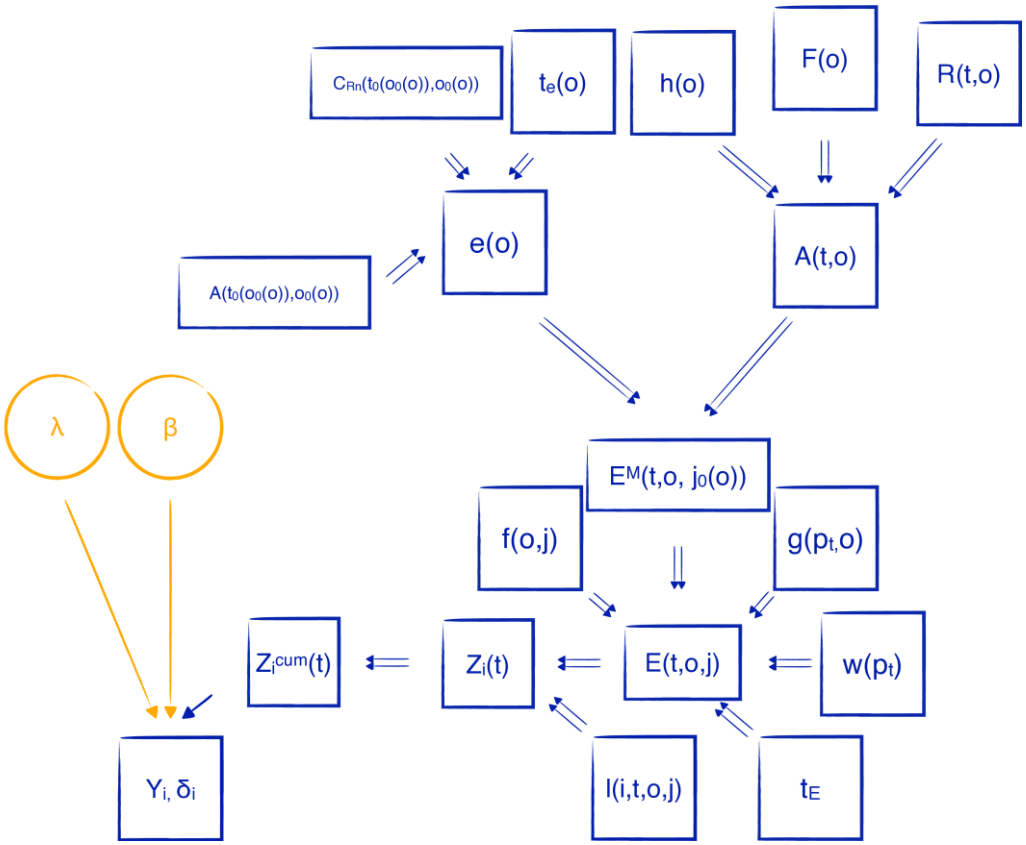

Figure 2.4: Exposure assessment in underground mining objects as well as in development objects in Thuringia in the first exposure assessment period.

For underground mining and development objects in Thuringia in the first exposure assessment period, the parameters are defined as follows:

- $f(o,j)$: activity weighting factor
- $w(p_t)$: working time factor
- $g(p_t,o)$: equilibrium factor
- Measurements of radon gas
  - $C_{Rn}(1955{,}903)$ and $C_{Rn}(1962{,}904)$: mean concentration measurements of radon gas for objects 903 BB Schmirchau in 1955 and for 904 BB Paitzdorf in 1962 were used to estimate the evaluation factors $e(o)$ for these two objects
- $t_e(o)$: multiplicative factor to determine the evaluation factor of non-reference objects as a function of the reference objects (Küchenhoff et al., 2018, p. 50; Lehmann et al., 1998, p. 67)
- t_E: multiplicative transfer factor (only for development objects)
- $A(1955{,}903)$ and $A(1962{,}904)$: evaluation area for objects 903 BB Schmirchau in 1955 and for 904 BB Paitzdorf in 1962 were used to estimate the evaluation factors $e(o)$ for these two objects
- $h(o)$: density of the bedrock
- $F(o)$: total shaft output
- $R(t,o)$: amount of uranium recovery

2.2.5 Measurement model M1b for objects in Thuringia

As shown in Table *2.1*, we assume the evaluation area $A(t,o)$ and the evaluation area of the respective reference object and year $A(t_0(o_0(o)),o_0(o))$ to be known without measurement error. Figure *2.5* shows the DAG for measurement model M1b assumed for underground mining and development objects in the first exposure assessment period in Thuringia. The measurement model

assumed for exploration objects in Thuringia is combined with the measurement model described in Section *2.5* for surface areas affiliated to mining objects.

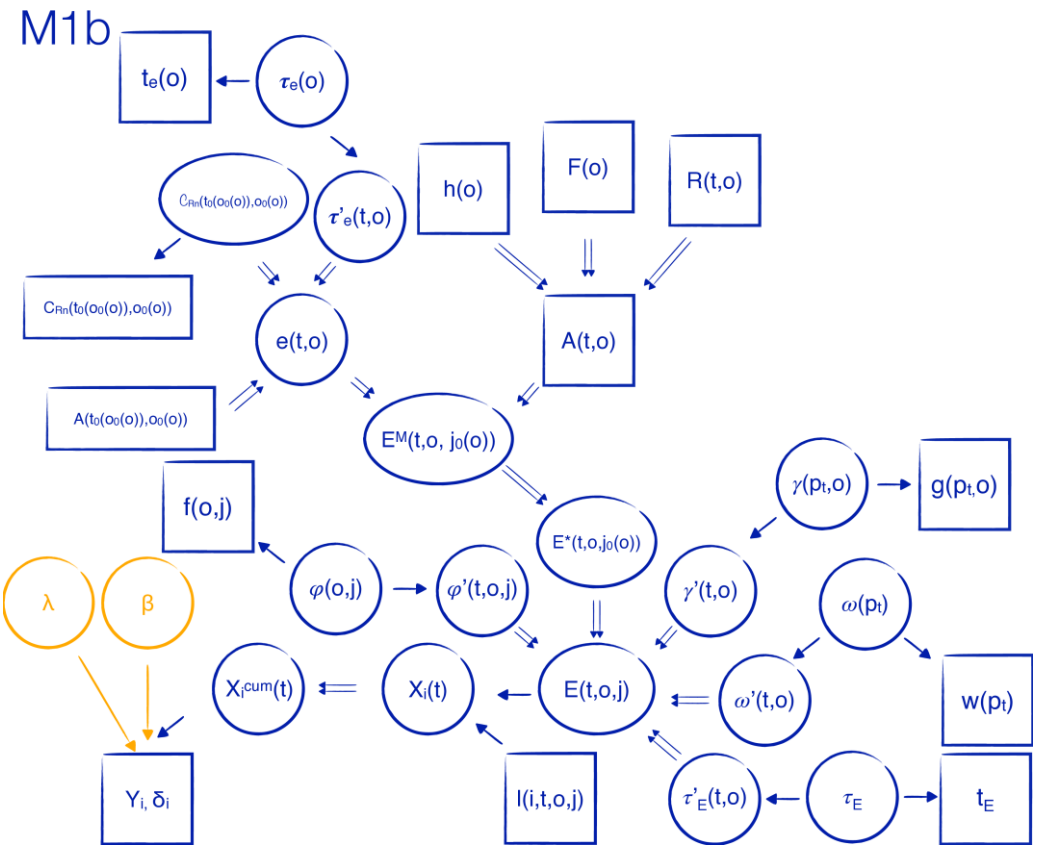

Figure 2.5: Hierarchical model combining a disease model with measurement model M1b to describe exposure uncertainty in underground mining objects as well as in development objects in Thuringia in the first exposure assessment period. Due to the limited space and for a clearer presentation, no measurement error variances are shown here.

For measurement model M1b we assume an additive classical measurement error for the mean radon gas concentration measurement $C_{Rn}(t_0(o_0(o)), o_0(o))$ and both a multiplicative classical and a multiplicative Berkson error component for the remaining uncertain quantities. Similar to measurement model M1a, the transfer factor t_E is only relevant for development objects in Thuringia.

$$
\begin{aligned}
C_{Rn}(t_0(o_0(o)), o_0(o)) &= \mathcal{C}_{Rn}(t_0(o_0(o)), o_0(o)) + U_{\mathcal{C},c}(t_0(o_0(o)), o_0(o)) \\
f(o,j) &= \varphi(o,j) \cdot U_{\varphi,c}(o,j) \\
\varphi'(t,o,j) &= \varphi(o,j) \cdot U_{\varphi',B}(t,o,j) \\
w(p_t) &= \omega(p_t) \cdot U_{\omega,c}(p_t) \\
\omega'(t,o) &= \omega(p_t) \cdot U_{\omega',B}(t,o) \\
g(p_t,o) &= \gamma(p_t,o) \cdot U_{\gamma,c}(p_t,o) \\
\gamma'(t,o) &= \gamma(p_t,o) \cdot U_{\gamma',B}(t,o) \\
t_e(o) &= \tau_e(o) \cdot U_{\tau_e,c}(o) \\
\tau_e{}'(t,o) &= \tau_e(o) \cdot U_{\tau_e',B}(t,o) \\
t_E &= \tau_E \cdot U_{\tau_E,c} \\
\tau_E{}'(t,o) &= \tau_E \cdot U_{\tau_E',B}(t,o)
\end{aligned}
$$

In summary, while the observed exposure $Z_i(t,o,j)$ for miner i in year t, object o and with activity j in the first exposure assessment period for underground mining objects in Thuringia was calculated as

$$Z_i(t,o,j) = \left(\frac{C_{Rn}(t_0(o_0(o)),o_0(o))}{A(t_0(o_0(o)),o_0(o))} \cdot t_e(o) \cdot A(t,o)\right) \cdot t_E \cdot g(p_t,o) \cdot w(p_t) \cdot f(o,j) \cdot 12 \cdot l(i,t,o,j),$$

the true value $X_i(t,o,j)$ for this exposure depends on the unknown quantities defined in the previous equations in the following way:

$$X_i(t,o,j) = \left(\frac{\mathcal{C}_{Rn}(t_0(o_0(o)),o_0(o))}{A(t_0(o_0(o)),o_0(o))} \cdot \tau_e{}'(t,o) \cdot A(t,o)\right) \cdot \tau_E{}'(t,o) \cdot \gamma'(t,o) \cdot \omega'(t,o) \cdot \varphi'(t,o,j) \cdot 12 \cdot l(i,t,o,j) + U_{E,B}(i,t)$$

where $l(i,t,o,j)$ reflects the individual working history of miner i.

2.3 Uncertainties in the exposure assessment based on radon gas concentration measurements in underground mining objects (1955/56 - 1965 in Saxony, 1955/56 - 1974 in Thuringia)

2.3.1 Exposure assessment

The first measurements of radon gas concentration in the different mines started in 1955, with the number of measurements changing over the years (see Küchenhoff et al. (2018, pp. 34, 37)). The length of this second period of radon exposure assessment in the Wismut cohort differs between the regions since the radon progeny concentration measurements in Thuringia were deficient until 1974 (Lehmann et al., 1998, p. 133). The second exposure assessment period with 18% PPY hence starts around 1955/1956 and lasts until 1965 for Saxonian objects and until 1974 for Thuringian objects.

As described on page 52 in Küchenhoff et al. (2018), the mean radon gas concentration $C_{Rn}(t,o) \cdot 12$ was used to approximate the annual exposure to radon gas for the reference activity of a hewer with 2000 working hours per year. In order to transform this annual exposure to radon gas for 2000 working hours to the annual exposure of radon progeny for the actual number of working hours, it is multiplied with the equilibrium factor $g(p_t,o)$ and the working time factor $w(p_t)$. For those workers with other activities than the reference activity, the activity weighting factor $f(o,j)$ is applied. Thus, in the years 1954/55 - 1965 in Saxony and 1955/56 - 1974 in Thuringia, exposure to radon progeny for year t, object o and a worker conducting activity j was estimated according to the following formula:

$$E(t,o,j) = C_{Rn}(t,o) \cdot 12 \cdot f(o,j) \cdot w(p_t) \cdot g(p_t,o)$$

Figure *2.6* illustrates the DAG for the radon exposure assessment in the second exposure assessment period. Similar to the exposure assessment in the first exposure assessment period, the transfer factor t_E depicted in Figure *2.6* is set to $t_E = 1$ for objects in the second exposure assessment period other than development objects.

In the second exposure assessment period, the parameters are defined as follows:

- $C_{Rn}(t,o)$: mean concentration of the radon gas measurements
- $f(o,j)$: activity weighting factor
- $w(p_t)$: working time factor
- $g(p_t,o)$: equilibrium factor
- t_E: multiplicative transfer factor (only for development objects)

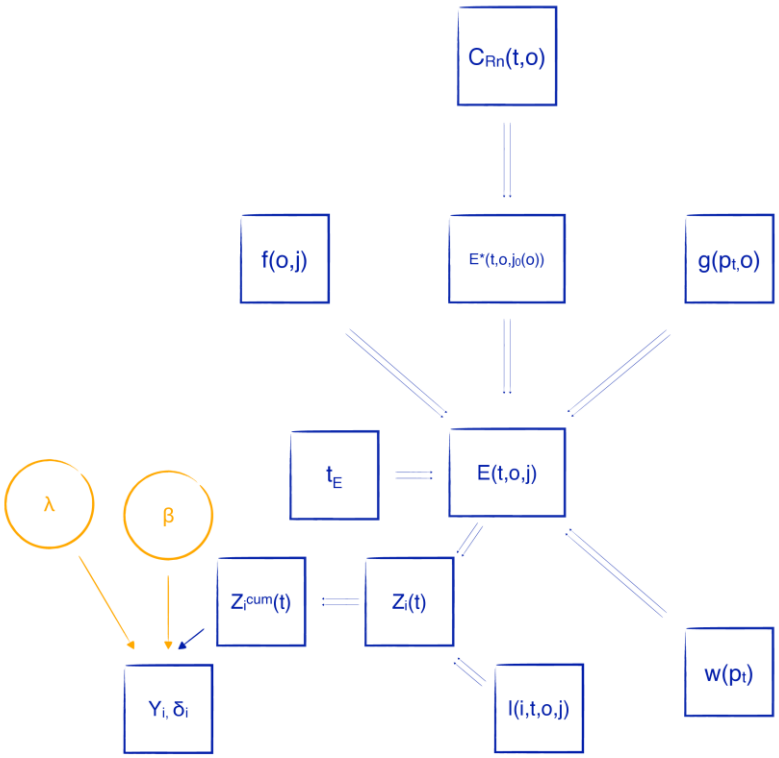

Figure 2.6: Exposure assessment in underground mining objects in Saxony and Thuringia as well as in development objects in the second exposure assessment period.

2.3.2 Measurement model M2

Figure *2.7* shows the DAG for measurement model M2 assumed for the second exposure assessment period.

For the mean radon gas concentration we assume an additive classical measurement error and for the remaining uncertain quantities both, a multiplicative classical and a multiplicative Berkson error component.
Similar to measurement models M1a and M1b, the transfer factor t_E is only relevant for development objects in the second exposure assessment period.

$$\begin{aligned} C_{Rn}(t,o) &= \mathcal{C}_{Rn}(t,o) + U_{\mathcal{C},c}(t,o) \\ f(o,j) &= \varphi(o,j) \cdot U_{\varphi,c}(o,j) \\ \varphi'(t,o,j) &= \varphi(o,j) \cdot U_{\varphi',B}(t,o,j) \\ w(p_t) &= \omega(p_t) \cdot U_{\omega,c}(p_t) \\ \omega'(t,o) &= \omega(p_t) \cdot U_{\omega',B}(t,o) \\ g(p_t,o) &= \gamma(p_t,o) \cdot U_{\gamma,c}(p_t,o) \\ \gamma'(t,o) &= \gamma(p_t,o) \cdot U_{\gamma',B}(t,o) \\ t_E &= \tau_E \cdot U_{\tau_E,c} \\ \tau_E{}'(t,o) &= \tau_E \cdot U_{\tau_E',B}(t,o) \end{aligned}$$

In summary, while the observed exposure $Z_i(t,o,j)$ for miner i in year t, object o and with activity j in underground mining objects in the second exposure assessment period was calculated as

$$Z_i(t,o,j) = C_{Rn}(t,o) \cdot 12 \cdot t_E \cdot g(p_t,o) \cdot w(p_t) \cdot f(o,j) \cdot l(i,t,o,j),$$

the true value $X_i(t,o,j)$ for this exposure depends on the unknown quantities defined in the previous equations in the following way:

$$X_i(t,o,j) \quad = \mathcal{C}_{Rn}(t,o) \cdot 12 \cdot \tau_E{}'(t,o) \cdot \gamma'(t,o) \cdot \omega'(t,o) \cdot \varphi'(t,o,j) \cdot l(i,t,o,j) + U_{E,B}(i,t)$$

where $l(i,t,o,j)$ reflects the individual working history of miner i.

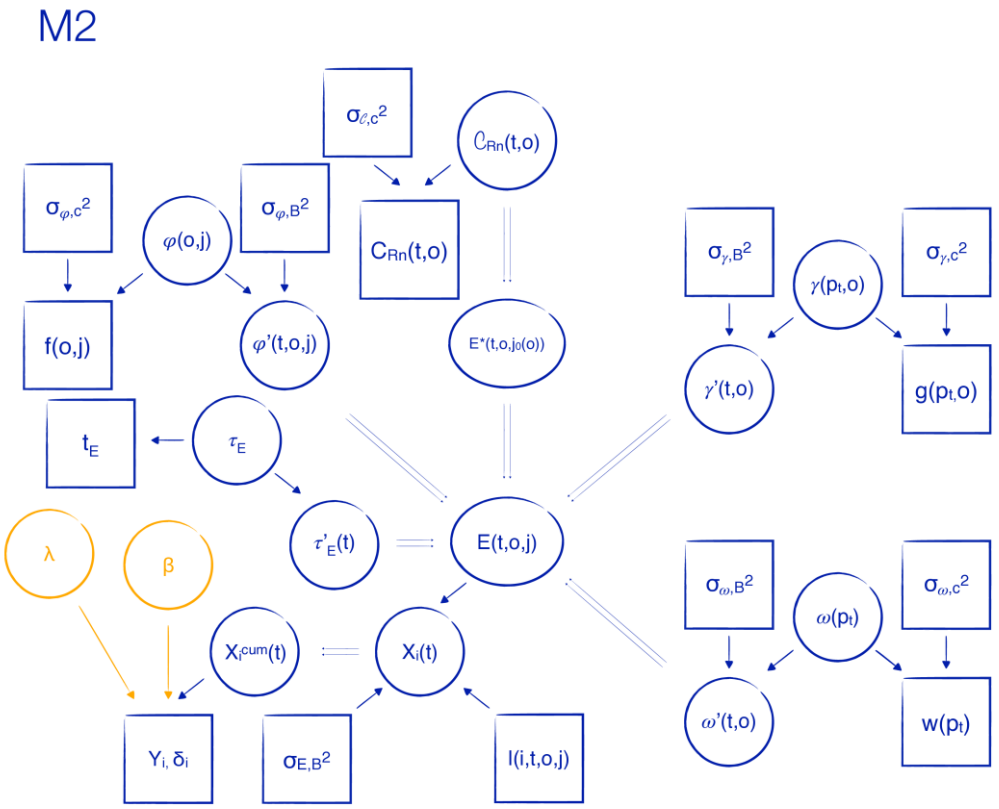

Figure 2.7: Hierarchical model combining a disease model with measurement model M2 to describe exposure uncertainty in underground mining objects in Saxony and Thuringia as well as in development objects in the second exposure assessment period. Due to the limited space and for a clearer presentation, no measurement error variances for t_E and $\tau'_E(t)$ are shown here.

2.4 Uncertainties in the exposure assessment based on radon progeny concentration measurements in underground mining objects (1966 - 1990 in Saxony, 1975 - 1990 in Thuringia)

2.4.1 Exposure assessment

In the third period of radon exposure assessment in the Wismut cohort (PPY: 25%), the radon progeny concentration in the different mines is measured directly. This third period lasts from 1966 for objects in Saxony and from 1975 for objects in Thuringia until the end of the uranium mining in the Wismut in 1990.

As described in Küchenhoff et al. (2018, p. 53), the mean radon progeny concentration $C_{RDP}(t,o) \cdot 12$ was used to approximate the annual exposure to radon progeny for the reference activity of a hewer with 2000 working hours per year. Since these measurements only partially account for disruptions and deficits of the ventilation system, the ventilation correction factor $c(o)$ corrects for such disruptions. By multiplying with the working time factor, the annual exposure of radon progeny for the actual number of working hours is obtained. Again, to obtain the exposure for workers conducting other activities than the reference activity, the activity weighting factor $f(o,j)$ is applied. In summary, in the years 1966 - 1990 in Saxony and 1975 - 1990 in Thuringia, exposure to radon progeny for year t, object o and a worker conducting activity j was estimated according to the following formula:

$$E(t,o,j) \quad = C_{RDP}(t,o) \cdot 12 \cdot f(o,j) \cdot w(p_t) \cdot c(o).$$

The exposure assessment in the third exposure assessment period is shown in the DAG in Figure *2.8*. As in the exposure assessment in the first and second exposure assessment period, the transfer factor t_E in Figure *2.8* is only relevant for development objects in the third exposure assessment period and is set to $t_E = 1$ for the remaining objects.

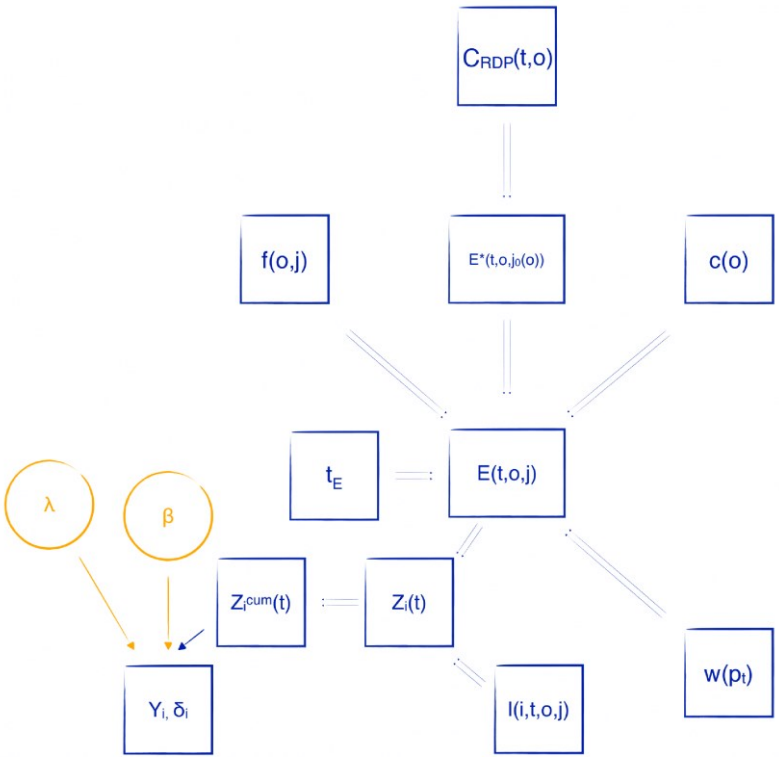

Figure 2.8: Exposure assessment in underground mining objects in Saxony and Thuringia as well as in development objects in the third exposure assessment period.

In the third exposure assessment period, the parameters are defined as follows:

- $C_{RDP}(t, o)$: mean concentration of the radon progeny measurements
- $f(o, j)$: activity weighting factor
- $w(p_t)$: working time factor
- $c(o)$: ventilation correction factor
- t_E: multiplicative transfer factor (only for development objects)

2.4.2 Measurement model M3

Figure *2.9* shows the DAG for the measurement model M3 assumed for the third exposure assessment period.

Similar to the second exposure assessment period, for the third exposure assessment period we assume an additive classical measurement error for the mean radon progeny concentration and a multiplicative classical as well as a multiplicative Berkson measurement error component for the remaining uncertain quantities in the exposure assessment. Similar to measurement models M1a, M1b and M2, the transfer factor t_E is only relevant for development objects.

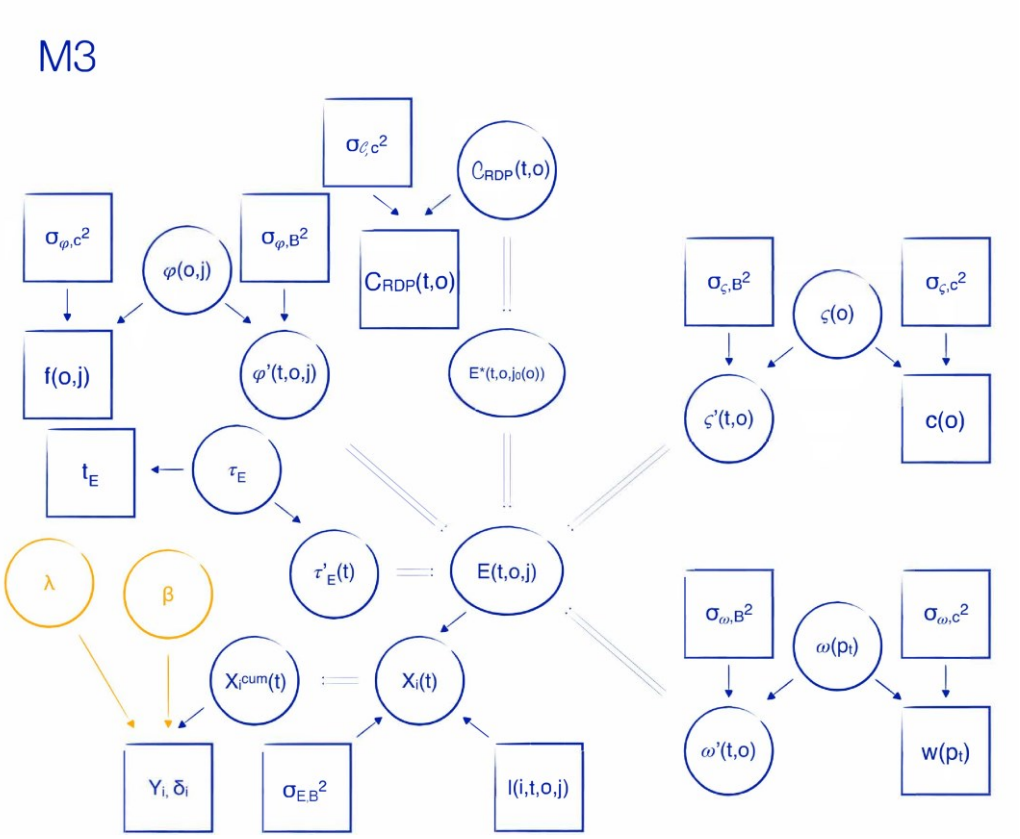

Figure 2.9: Hierarchical model combining a disease model with measurement model M3 to describe exposure uncertainty in underground mining objects in Saxony and Thuringia as well as in development objects in the third exposure assessment period. Due to the limited space and for a clearer presentation, no measurement error variances for t_E and $\tau'_E(t)$ are shown here.

$$
\begin{aligned}
C_{RDP}(t,o) &= \mathcal{C}_{RDP}(t,o) + U_{\mathcal{C},c}(t,o) \\
f(o,j) &= \varphi(o,j) \cdot U_{\varphi,c}(o,j) \\
\varphi'(t,o,j) &= \varphi(o,j) \cdot U_{\varphi',B}(t,o,j) \\
w(p_t) &= \omega(p_t) \cdot U_{\omega,c}(p_t) \\
\omega'(t,o) &= \omega(p_t) \cdot U_{\omega',B}(t,o) \\
c(o) &= \varsigma(o) \cdot U_{\varsigma,c}(o) \\
\varsigma'(t,o) &= \varsigma(o) \cdot U_{\varsigma',B}(t,o) \\
t_E &= \tau_E \cdot U_{\tau_E,c} \\
\tau_E'(t,o) &= \tau_E \cdot U_{\tau_E',B}(t,o)
\end{aligned}
$$

In summary, while the observed exposure $Z_i(t,o,j)$ for miner i in year t, object o and with activity j in underground mining objects in the third exposure assessment period was calculated as

$$Z_i(t,o,j) = C_{RDP}(t,o) \cdot 12 \cdot t_E \cdot c(o) \cdot w(p_t) \cdot f(o,j) \cdot l(i,t,o,j),$$

the true value $X_i(t,o,j)$ for this exposure depends on the unknown quantities defined in the previous equations in the following way:

$$X_i(t,o,j) = \mathcal{C}_{RDP}(t,o) \cdot 12 \cdot \tau_E'(t,o) \cdot \varsigma'(t,o) \cdot \omega'(t,o) \cdot \varphi'(t,o,j) \cdot l(i,t,o,j) + U_{E,B}(i,t)$$

where $l(i,t,o,j)$ reflects the individual working history of miner i.

2.5 Uncertainties in the exposure assessment in surface areas affiliated to mining objects and in exploration objects in Thuringia

2.5.1 Exposure assessment

The surface areas affiliated to mining objects (PPY: 21%) were used as facilities for the handling and transport of waste rock and ore (Küchenhoff et al., 2018, p. 55). Since the workers in these surface areas were exposed to high dust concentrations, they were also included in the JEM (Lehmann et al., 1998, p. 107).

The mined uranium concentrations of the different objects as well as the exposure during the uranium ore loading was considered for the radon progeny exposure $E(p_t, o)$. According to Lehmann et al. (1998), the value of this radon progeny exposure for surface areas affiliated to mining objects $E(p_t, o)$ was determined by experts. Here, radon progeny concentration does not depend on t, but rather on p_t, because these values were often estimated jointly for a period of time p_t, creating a dependence structure for several exposure years that is represented through p_t. Thus, all workers with reference activity in one respective object o are assigned the same value of $E(p_t, o)$ for all years belonging to p_t. The respective determined radon progeny exposure describes the exposure of a worker in ore milling, which is the reference activity for workers in surface areas affiliated to mining objects (Küchenhoff et al., 2018, p. 55). In order to obtain the annual radon exposure for other workers than those in ore milling, the radon progeny exposure has to be multiplied with an activity weighting factor $f(o, j)$.

$$E(p_t, o, j) = E(p_t, o) \cdot f(o, j)$$

Note that the exposure assessment in exploration objects in Thuringia corresponds to the exposure assessment in surface areas affiliated to mining objects (Lehmann et al., 1998, p. 140).

Figure *2.10* shows the DAG for the exposure assessment in surface areas affiliated to mining objects and in exploration objects in Thuringia.

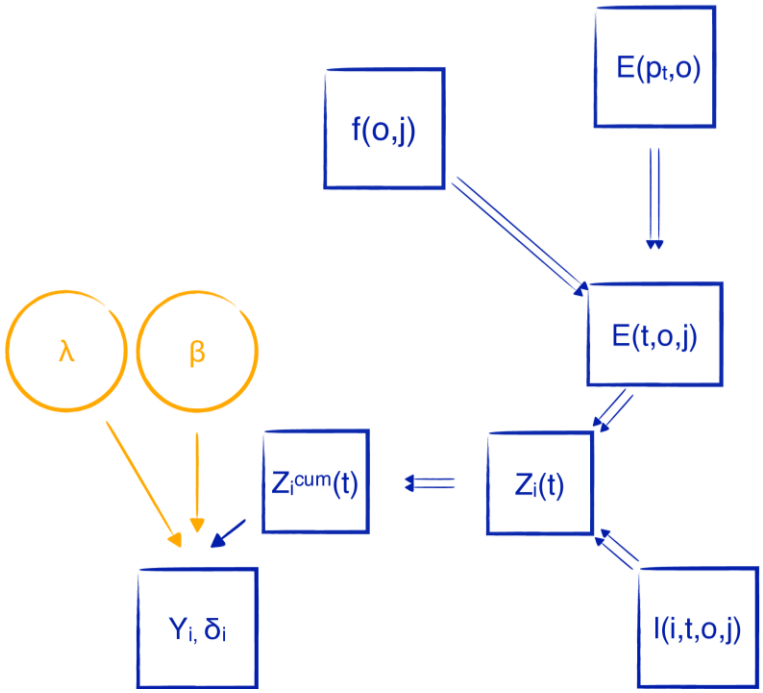

Figure 2.10: Exposure assessment in surface areas affiliated to mining objects and in exploration objects in Thuringia.

For the exposure assessment in surface areas affiliated to mining objects and in exploration objects in Thuringia, the parameters are defined as follows:

- $E(p_t, o)$: radon progeny exposure
- $f(o, j)$: activity weighting factor

2.5.2 Measurement model M4

Since the assessment of radon exposure is very similar in surface areas affiliated to mining objects and in exploration objects in Thuringia (see Section *2.2.4*), we use the same measurement model M4 for both. Note though, that the activity weighting factor $f(o, j)$ is different for exploration objects in Thuringia than for surface areas affiliated to mining objects. The reference activity for exploration objects is a hewer, whereas the reference activity for surface areas is a worker in ore milling.

Figure *2.11* shows the DAG for measurement model M4 assumed for surface areas affiliated to mining objects and for exploration objects in Thuringia.

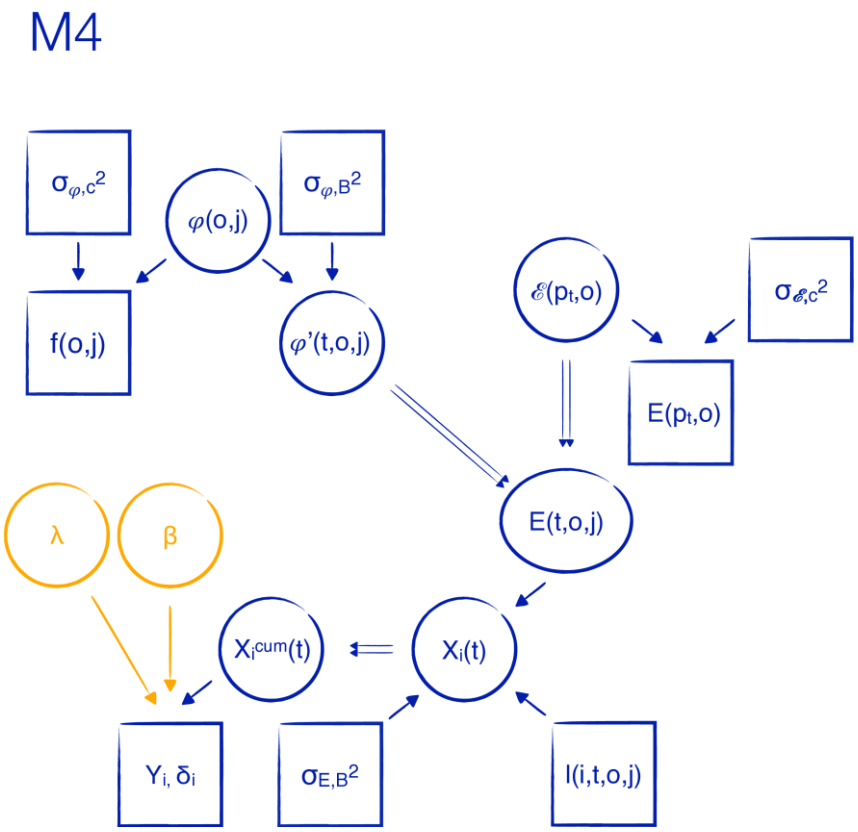

Figure 2.11: Hierarchical model combining a disease model with measurement model M4 to describe exposure uncertainty in surface areas affiliated to mining objects and in exploration objects in Thuringia.

Here we assume a multiplicative classical measurement error for the radon progeny exposure $E(p_t, o)$ and both a multiplicative classical and a multiplicative Berkson error component for the activity weighting factor $f(o, j)$.

$$\begin{aligned} E(p_t, o) &= \mathcal{E}(p_t, o) \cdot U_{E,c}(p_t, o) \\ f(o, j) &= \varphi(o, j) \cdot U_{\varphi,c}(o, j) \\ \varphi'(t, o, j) &= \varphi(o, j) \cdot U_{\varphi',B}(t, o, j) \end{aligned}$$

In summary, while the observed exposure $Z_i(t, o, j)$ for miner i in year t, object o and with activity j in surface areas affiliated to mining objects and in exploration objects in Thuringia was calculated as

$$Z_i(t, o, j) = E(p_t, o) \cdot f(o, j) \cdot I(i, t, o, j),$$

the true value $X_i(t, o, j)$ for this exposure depends on the unknown quantities defined in the previous equations in the following way:

$$X_i(t,o,j) \quad = \mathcal{E}(p_t,o) \cdot \varphi'(t,o,j) \cdot l(i,t,o,j) + U_{E,B}(i,t)$$

where $l(i,t,o,j)$ reflects the individual working history of miner i.

2.6 Uncertainties in the exposure assessment in processing companies

2.6.1 Exposure assessment

As described in Küchenhoff et al. (2018, pp. 14, 27, 28), processing companies (PPY: 7%) had the task to process the mined ore. They can be divided into two different exposure assessment periods for processing (first period: 2% PPY; second period: 5% PPY) and into three different types: processing facilities, RAS and RAF facilities (RAS: radiometric automatic gradation; RAF: radiometric processing factory) and collieries (Lehmann et al., 1994, 1998).

Contrary to mining activities, the radon exposure for workers in the processing companies was assessed on the basis of processing stages (Küchenhoff et al., 2018, p. 60). The activity specific radon exposure for 2000 working hours was determined by either assigning the activity to a processing stage or by the time-weighted mean exposure of several processing stages. Since the assignment of activities to processing stages is not directly evident from the available literature and since the reconstruction of this assignment was not feasible within this project, we will directly use the activity specific radon exposure for 2000 working hours, which is depicted in Lehmann et al. (1998, pp. 358–369, 394–405). Since we use this activity specific exposure and not the exposure in the processing stages, no activity weighting factor is needed for most workers in processing companies. An exception are workers in the auxiliaries, managing staff and other personnel with frequently changing work places, whose exposure was determined as the weighted exposure of a colliery worker. Only for these activities the activity weighting factor $f(j)$ is applied (Lehmann et al., 1998, pp. 483–484).

The radon exposure assessment for processing companies is different for the two exposure assessment periods for processing. In the first exposure assessment period for processing, there were only very few single radon gas measurements, so that the annual radon progeny exposure $E(o_0,j)$ had to be estimated retrospectively by experts (Lehmann et al., 1994). In the second exposure assessment period for processing, the exposure assessment was based on radon gas concentration measurements, which were multiplied with an equilibrium factor $g = 0.4$ to obtain radon progeny concentrations (Küchenhoff et al., 2018, p. 60). Since there were only enough measurements for some objects, these objects were chosen as the reference objects. For processing facilities, the objects 101 Crossen and 102 Seelingstädt were the reference objects, the RAF facility of object 009 Aue (shaft 371) was the reference object for RAS and RAF facilities and the colliery 050 for collieries (Lehmann et al., 1994, Chapter 4). For non-reference objects, the annual exposure for 2000 working hours is estimated by weighting the radon progeny exposure of the respective reference object $E^*(p(t,o),o_0,j)$ with an object weighting factor $z(o,j)$ (Küchenhoff et al., 2018, p. 62). Finally, to obtain the radon exposure for the actual annual working time, the working time factor $w(p_t,o)$ is needed, which varies depending on the time period and the type of processing object (Lehmann et al., 1994, p. 89).

Thus, in summary the annual exposure to radon progeny in processing companies in the first exposure assessment period for processing was estimated according to the following formula:

$$E(t,o,j) = E(o_0,j) \cdot w(p_t,o) \cdot z(o,j) \cdot f(j)$$

and for processing companies in the second exposure assessment period for processing according to the formula:

$$E(t,o,j) = E^*(p(t,o),o_0,j) \cdot g \cdot w(p_t,o) \cdot z(o,j) \cdot f(j)$$

The exposure assessment in the first exposure assessment period for processing is illustrated with the DAG in Figure *2.12* and Figure *2.13* shows the DAG for the exposure assessment in the second exposure assessment period for processing.

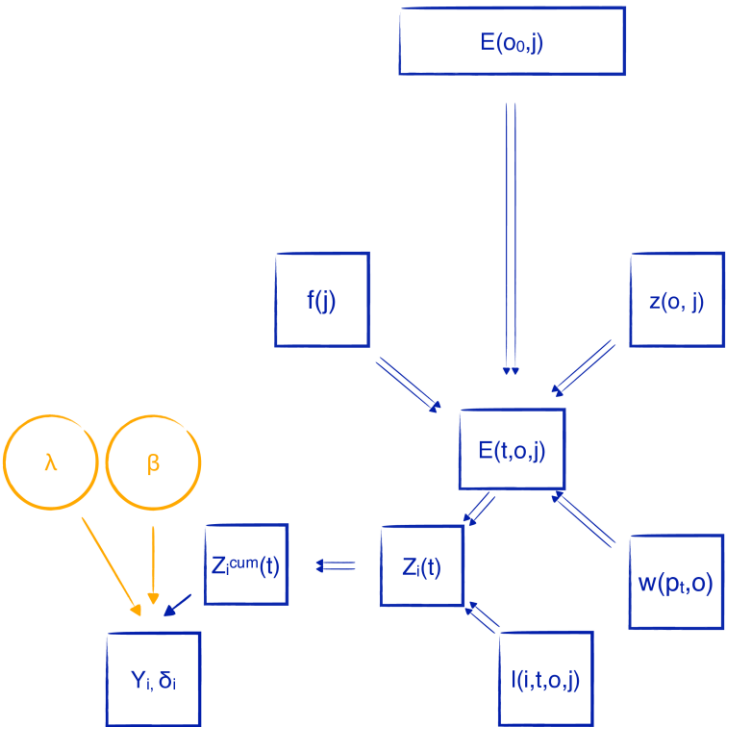

Figure 2.12: Exposure assessment in processing companies in the first exposure assessment period for processing.

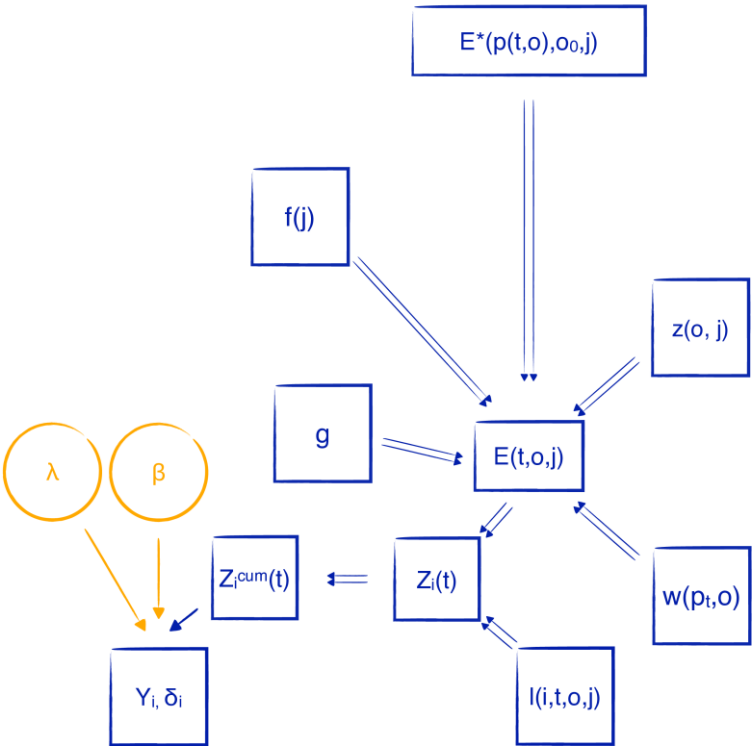

Figure 2.13: Exposure assessment in processing companies in the second exposure assessment period for processing.

For the exposure assessment in processing companies, the parameters are defined as follows:

- $E(o_0, j)$: annual exposure to radon progeny of a worker in a reference object with 2000 working hours in the first exposure assessment period for processing
- $E^*(p(t,o), o_0, j)$: annual exposure to radon gas of a worker in a reference object with 2000 working hours in the second exposure assessment period for processing
- g: equilibrium factor (only relevant for the second exposure assessment period for processing)
- $w(p_t, o)$: working time factor
- $z(o, j)$: object weighting factor
- $f(j)$: activity weighting factor

2.6.2 Measurement models M5a and M5b

We distinguish the measurement models for processing companies into one model for the first exposure assessment period for processing (M5a) and one model for the second exposure assessment period for processing (M5b). Figures *2.14* and *2.15* show the DAGs for the measurement models for processing companies.

Here we assume a multiplicative classical error for $E(o_0, j)$ (first exposure assessment period for processing) or $E^*(p(t,o), o_0, j)$ (second exposure assessment period for processing) and multiplicative classical and Berkson error components for the activity, object weighting, working time and equilibrium factors.

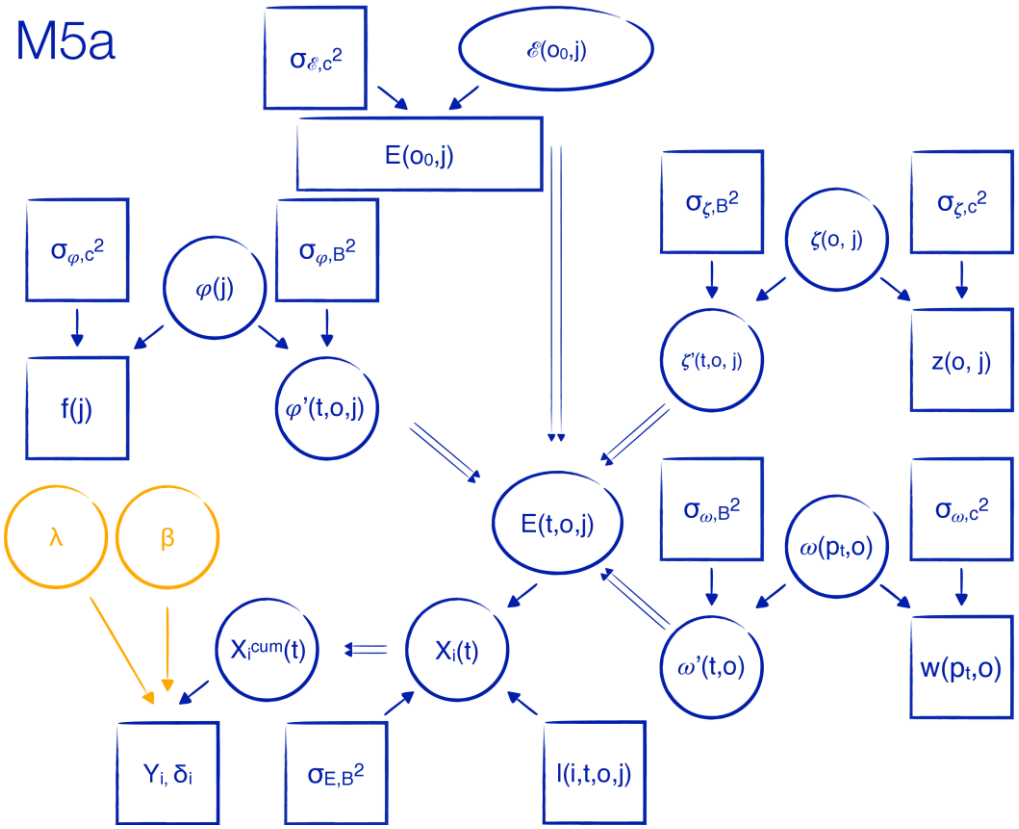

Figure 2.14: Hierarchical model combining a disease model with measurement model M5a to describe exposure uncertainty in processing companies in the first exposure assessment period for processing companies.

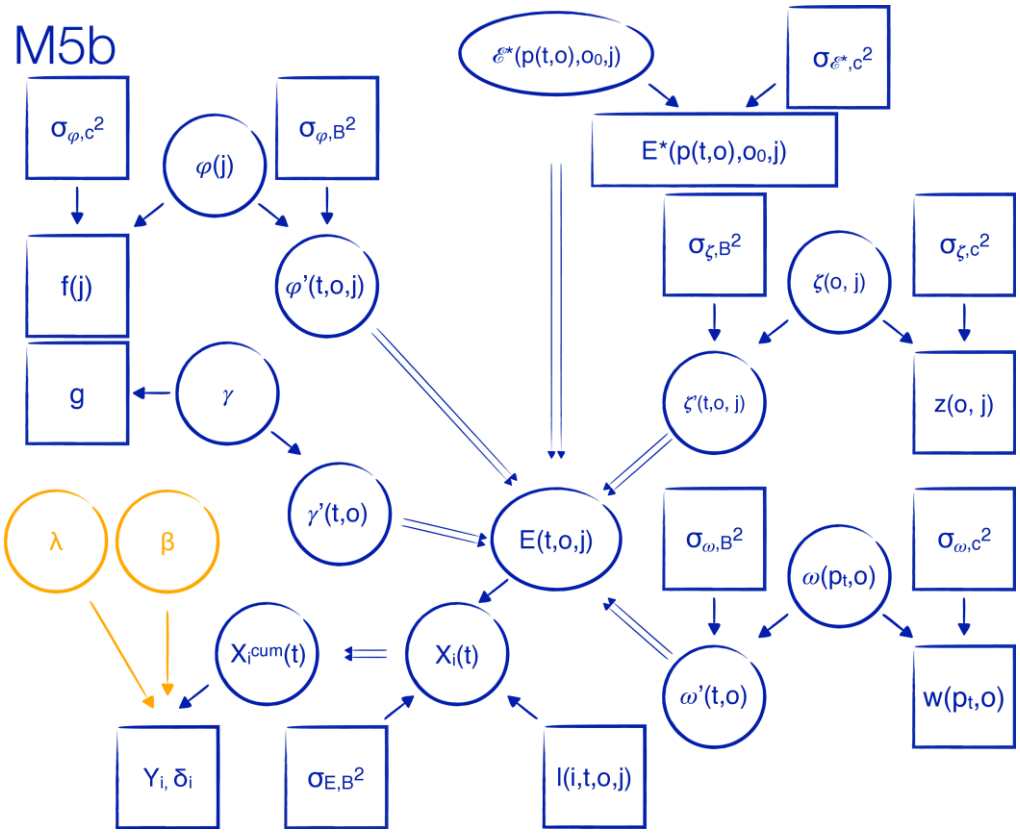

Figure 2.15: Hierarchical model combining a disease model with measurement model M5b to describe exposure uncertainty in processing companies in the second exposure assessment period for processing companies. Due to the limited space and for a clearer presentation, no measurement error variances for g and $\gamma'(t,o)$ are shown here.

First exposure assessment period for processing:

$$E(o_0,j) \quad = \mathcal{E}(o_0,j) \cdot U_{\mathcal{E},c}(o_0,j)$$

Second exposure assessment period for processing:

$$E^*(p(t,o),o_0,j) \quad = \mathcal{E}^*(p(t,o),o_0,j) \cdot U_{\mathcal{E}^*,c}(p(t,o),o_0,j)$$

Second exposure assessment period for processing:

$$\begin{aligned} g \quad &= \gamma \cdot U_{\gamma,c} \\ \gamma'(t,o) \quad &= \gamma \cdot U_{\gamma',B}(t,o) \end{aligned}$$

First and second exposure assessment period for processing:

$$\begin{aligned} f(j) \quad &= \varphi(j) \cdot U_{\varphi,c}(j) \\ \varphi'(t,o,j) \quad &= \varphi(j) \cdot U_{\varphi',B}(t,o,j) \\ z(o,j) \quad &= \zeta(o,j) \cdot U_{\zeta,c}(o,j) \\ \zeta'(t,o,j) \quad &= \zeta(o,j) \cdot U_{\zeta',B}(t,o,j) \\ w(p_t,o) \quad &= \omega(p_t,o) \cdot U_{\omega,c}(p_t,o) \\ \omega'(t,o) \quad &= \omega(p_t,o) \cdot U_{\omega',B}(t,o) \end{aligned}$$

In summary, the observed exposure $Z_i(t,o,j)$ for miner i in year t, object o and with activity j in processing companies was calculated as:

First exposure assessment period for processing:

$$Z_i(t,o,j) \quad = E(o_0,j) \cdot w(p_t,o) \cdot z(o,j) \cdot f(j) \cdot l(i,t,o,j)$$

Second exposure assessment period for processing:

$$Z_i(t,o,j) \quad = E^*(p(t,o),o_0,j) \cdot g \cdot w(p_t,o) \cdot z(o,j) \cdot f(j) \cdot l(i,t,o,j)$$

The true values $X_i(t,o,j)$ for these exposures depend on the unknown quantities defined in the previous equations in the following way:

First exposure assessment period for processing:

$$X_i(t,o,j) \quad = \mathcal{E}(o_0,j) \cdot \omega'(t,o) \cdot \zeta'(t,o,j) \cdot \varphi'(t,o,j) \cdot l(i,t,o,j) + U_{E,B}(i,t)$$

Second exposure assessment period for processing:

$$X_i(t,o,j) \quad = \mathcal{E}^*(p(t,o),o_0,j) \cdot \gamma'(t,o) \cdot \omega'(t,o) \cdot \zeta'(t,o,j) \cdot \varphi'(t,o,j) \cdot l(i,t,o,j) + U_{E,B}(i,t)$$

where $l(i,t,o,j)$ reflects the individual working history of miner i.

2.7 Uncertainties in the exposure assessment in open pit mining objects

2.7.1 Exposure assessment

During the operation time of the Wismut there were no radon gas or radon progeny concentration measurements in open pit mining objects (PPY: 1%). Therefore, all exposure values for open pit mining objects are retrospective expert estimations which are based on measurements of radon gas in the object 300 Lichtenberg in 1994/1995 (Küchenhoff et al., 2018). These measurements were performed at ground level and in a depth of 130m (Lehmann et al., 1998, p. 146).

Following Küchenhoff et al. (2018), the annual exposure to radon gas for 2000 working hours per year in open pit mining objects can be estimated by the sum of the basic radon gas concentration at ground level without mining activity and the radon gas concentration in the depth. The mean of the measurements at ground level in object 300 Lichtenberg in 1994/1995, $C_{Rn,0}(1994/1995{,}300) = 30$ Bq/m^3 was used as the basic radon gas concentration for all open pit mining objects. The radon exposure in the depth was reconstructed as follows. Firstly, the difference between the means of the radon gas concentration measurements in object 300 Lichtenberg in 1994/1995 in a depth of 130m ($C_{Rn,130}(1994/1995{,}300) = 80$ Bq/m^3) and the measurements at ground level was used to obtain the concentration without mining activity in a depth of 130m. Secondly, this difference was multiplied with the ratio between the depth $d(t,o)$ in year t of object o and the depth of 130m. At last, the resulting radon gas concentration in the depth without mining activity was multiplied with the evaluation factor in open pit mining objects, $e(p_t,o)$, which evaluates the conditions in the object and year of interest compared to the reference object 300 Lichtenberg in the reference year 1994/1995. In contrast to the evaluation factor for underground mining objects, the evaluation factor for open pit mining objects consists of six sub-factors:

$$e(p_t,o) = e_1 \cdot e_2(p_t,o) \cdot e_3(p_t,o) \cdot e_4(p_t,o) \cdot e_5(p_t,o) \cdot e_6(o)$$

These sub-factors evaluate the following aspects of mining conditions:

- e_1: diggings
- $e_2(p_t,o)$: production
- $e_3(p_t,o)$: weather exchange
- $e_4(p_t,o)$: fire events
- $e_5(p_t,o)$: underground blowing ventilation
- $e_6(o)$: uranium mineralization

Table 2.2 shows the values of the sub-factors for the different open pit mining objects and years.

Table 2.2: Evaluation sub-factors for evaluating the conditions in the open pit mining objects over time, adapted from Lehmann et al. (1998, Tables 5.4.2.2.3 and 5.4.2.3.1)

object	object name	year	e_1 (diggings)	e_2 (production)	e_3 (weather exchange)	e_4 (fire events)	e_5 (underground blowing ventilation)	e_6 (uranium mineralization)
300	Lichtenberg	1959	1.2	20	1	1.25	1	1
300	Lichtenberg	1960	1.2	25	1	1.25	1	1
300	Lichtenberg	1964	1.2	25	1	1.25	1	1
300	Lichtenberg	1968	1.2	25	1.3	1.25	1	1
300	Lichtenberg	1975	1.2	25	1.3	1.25	1.1	1
300	Lichtenberg	1977	1.2	5	1.3	1	1.1	1
301	Stolzenberg		1.2	20	1	1	1	1
302	Ronneburg/Raitzhain		1.2	15	1	1	1	1
303	Sorge		1.2	20	1	1	1	1
304	Gauern		1.2	15	1	1	1	1
306	Culmitzsch		1.2	25	1	1	1	1.2
307	Trünzig		1.2	15	1	1	1	1
308	Steinach		1.2	15	1	1	1	1
309	Erlau/Hirschbach		1.2	15	1	1	1	1

Finally, to obtain the annual exposure to radon progeny for a miner working in an open pit mining object, the annual exposure to radon gas for 2000 working hours per year has to be multiplied with the equilibrium factor g, the working time factor $w(p_t)$ and the activity weighting factor $f(p_t, j)$. To convert the resulting radon exposure in Bq/m^3 to WLM, the factor $\frac{12}{3700}$ is applied. Thus, the annual exposure to radon progeny for a miner working in an open pit mining object can be estimated as:

$$E(t,o,j) = \frac{12}{3700}\Big(C_{Rn,0}(1994/1995{,}300) + \left(C_{Rn,130}(1994/1995{,}300) - C_{Rn,0}(1994/1995{,}300)\right)\frac{d(t,o)}{130}\cdot e(p_t,o)\Big) \cdot g \cdot w(p_t) \cdot f(p_t,j)$$

Figure *2.16* illustrates the DAG for the exposure assessment in open pit mining objects.

Given that most evaluation sub-factors remained constant for many years and many objects and for the sake of simplicity, we combine e_1, $e_3(p_t,o)$, $e_4(p_t,o)$, $e_5(p_t,o)$ and $e_6(o)$ to a common sub-factor $e(p_t,o)$ and only consider production $e_2(p_t,o)$ as an independent sub-factor.
For the exposure assessment in open pit mining objects, the parameters are defined as follows:

- $f(p_t,j)$: activity weighting factor for open pit mining
- $w(p_t)$: working time factor
- g: equilibrium factor
- $e(p_t,o)$: evaluation factor in open pit mining objects combining the following sub-factors evaluating different aspects of mining conditions
  - e_1: diggings
  - $e_3(p_t,o)$: weather exchange
  - $e_4(p_t,o)$: fire events
  - $e_5(p_t,o)$: underground blowing ventilation
  - $e_6(o)$: uranium mineralization
- $e_2(p_t,o)$: sub-factor production
- Measurements of radon gas
  - $C_{Rn,0}(1994/1995{,}300)$: mean concentration at ground level without mining activity for object 300 Lichtenberg in 1994/1995

- $C_{Rn,130}(1994/1995,300)$: mean concentration of radon gas measurements in a depth of 130m for object 300 Lichtenberg in 1994/1995 without mining activity

- $d(t,o)$: depth

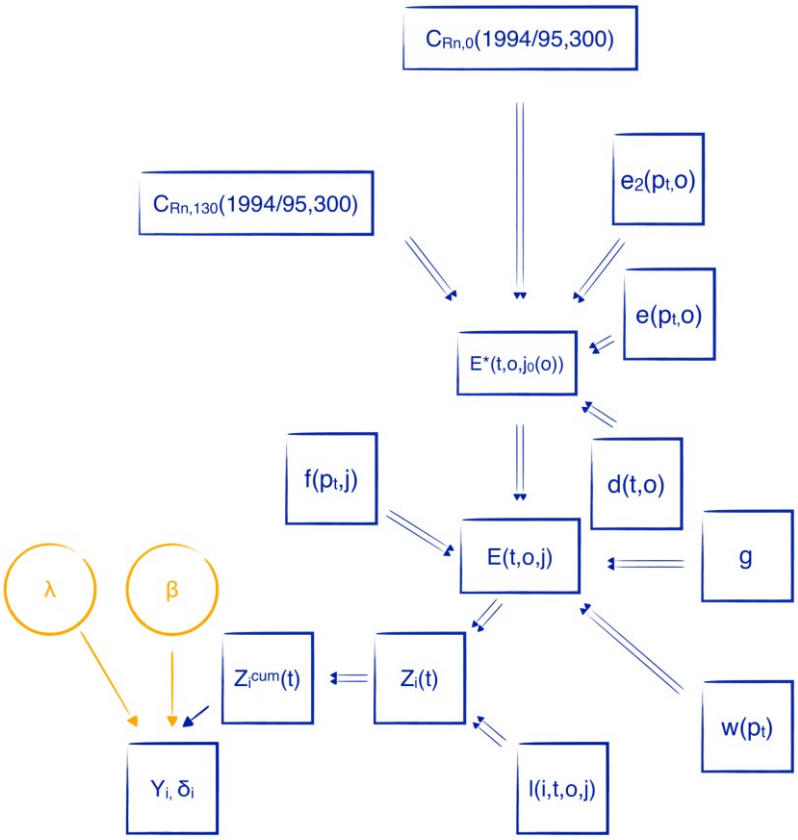

Figure 2.16: Exposure assessment in open pit mining objects.

2.7.2 Measurement model M6

Figure *2.17* shows the DAG for measurement model M6 for open pit mining objects.

We assume additive classical measurement errors for the mean concentration of radon gas measurements $C_{Rn,0}(1994/1995,300)$ and $C_{Rn,130}(1994/1995,300)$. As shown in Table *2.1*, the depth $d(t,o)$ is assumed to be measured with sufficient precision to consider it as known without measurement error. For all other uncertain quantities intervening in the exposure estimation for workers in open pit mining objects, we assume a multiplicative classical measurement error as well as a multiplicative Berkson error component.

$$\begin{aligned} C_{Rn,0}(1994/1995,300) &= \mathcal{C}_{Rn,0}(1994/1995,300) + U_{\mathcal{C},c}(1994/1995,300) \\ C_{Rn,130}(1994/1995,300) &= \mathcal{C}_{Rn,130}(1994/1995,300) + U_{\mathcal{C},c}(1994/1995,300) \end{aligned}$$

$$\begin{aligned} f(p_t,j) &= \varphi(p_t,j) \cdot U_{\varphi,c}(p_t,j) \\ \varphi'(t,o,j) &= \varphi(p_t,j) \cdot U_{\varphi',B}(t,o,j) \end{aligned}$$

$$\begin{aligned} w(p_t) &= \omega(p_t) \cdot U_{\omega,c}(p_t) \\ \omega'(t,o) &= \omega(p_t) \cdot U_{\omega',B}(t,o) \end{aligned}$$

$$\begin{aligned} g &= \gamma \cdot U_{\gamma,c} \\ \gamma'(t,o) &= \gamma \cdot U_{\gamma',B}(t,o) \end{aligned}$$

$$
\begin{aligned}
e(p_t, o) &= \epsilon(p_t, o) \cdot U_{\epsilon,c}(p_t, o) \\
\epsilon'(t, o) &= \epsilon(p_t, o) \cdot U_{\epsilon',B}(t, o) \\
e_2(p_t, o) &= \epsilon_2(p_t, o) \cdot U_{\epsilon_2,c}(p_t, o) \\
\epsilon'_2(t, o) &= \epsilon_2(p_t, o) \cdot U_{\epsilon'_2,B}(t, o)
\end{aligned}
$$

In summary, while the observed exposure $Z_i(t, o, j)$ for miner i in year t, object o and with activity j in open pit mining objects was calculated as

$$
\begin{aligned}
Z_i(t, o, j) &= \frac{12}{3700}\Big(C_{Rn,0}(1994/1995{,}300) \\
&\quad + (C_{Rn,130}(1994/1995{,}300) - C_{Rn,0}(1994/1995{,}300))\frac{d(t, o)}{130} \cdot e(p_t, o) \cdot e_2(p_t, o)\Big) \\
&\quad \cdot g \cdot w(p_t) \cdot f(p_t, j) \cdot l(i, t, o, j),
\end{aligned}
$$

the true value $X_i(t, o, j)$ for this exposure depends on the unknown quantities defined in the previous equations in the following way:

$$
\begin{aligned}
X_i(t, o, j) &= \frac{12}{3700}\Big(\mathcal{C}_{Rn,0}(1994/1995{,}300) \\
&\quad + (\mathcal{C}_{Rn,130}(1994/1995{,}300) - \mathcal{C}_{Rn,0}(1994/1995{,}300))\frac{d(t, o)}{130} \cdot \epsilon'(t, o) \cdot \epsilon_2'(t, o)\Big) \\
&\quad \cdot \gamma'(t, o) \cdot \omega'(t, o) \cdot \varphi'(t, o, j) \cdot l(i, t, o, j) + U_{E,B}(i, t)
\end{aligned}
$$

where $l(i, t, o, j)$ reflects the individual working history of miner i.

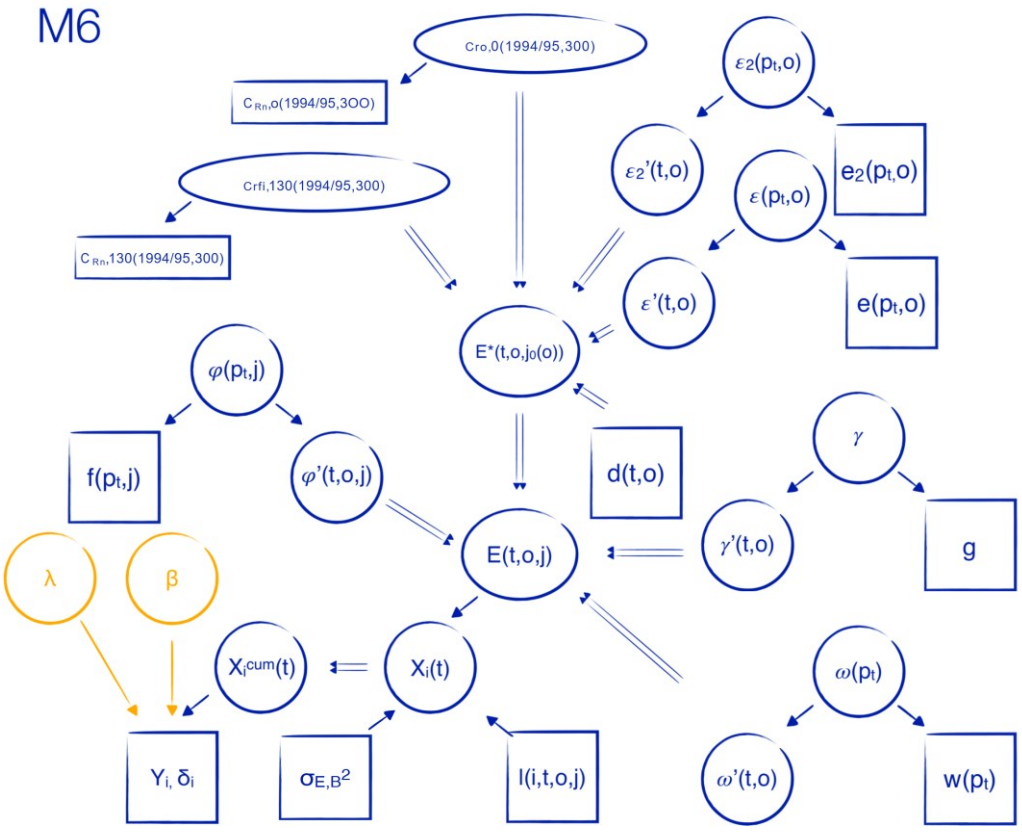

Figure 2.17: Hierarchical model combining a disease model with measurement model M6 to describe exposure uncertainty in open pit mining objects. Due to the limited space and for a clearer presentation, no measurement error variances are shown here.

2.8 Transfer of Job Exposure Matrix values

In addition to the quantities assumed to be measured with error, which were described for the eight measurement models in the last sections, a further transfer error might arise.

We will denote this transfer factor as t_{JEM}. It occurs whenever it is assumed for years with missing radon progeny exposure values in the JEM that the values for these missing years are the same as those of other years or objects for which there are exposure values. The transfer error t_{JEM} is added to the respective measurement models for all objects and years with continuation of radon exposure values from other years of the same object or the use of values from a different object to fill the gaps of years and objects with missing exposure values.
Similar to other uncertain quantities, we assume a multiplicative classical and a multiplicative Berkson error component for the transfer error for filled gaps in the exposure assessment.

$$\begin{aligned} t_{JEM} &= \tau_{JEM} \cdot U_{\tau_{JEM},c} \\ \tau'_{JEM}(t,o) &= \tau_{JEM} \cdot U_{\tau'_{JEM},B}(t,o) \end{aligned}$$

Note that due to space limitations, the transfer error for filled gaps in the exposure assessment t_{JEM} is not shown in any of the DAGs. Actually, it should be depicted in each DAG in the same way as t_E in the DAGs for M1a, M1b, M2 and M3, so that the double arrow from $\tau'_{JEM}(t,o)$ points to $E(t,o,j)$.

2.9 Characteristics of exposure uncertainty in the Wismut cohort

Küchenhoff et al. (2018, p. 76) identify three major types of measurement error in radiation exposure assessment in the Wimut cohort: generalization error, assignment error and estimation error. The authors give a detailed description of the different sources of uncertainty in Küchenhoff et al. (2018, pp. 78–88). The resulting classification is summarized in Table 2.3. Küchenhoff et al. (2018, p. 114) categorize these sources of uncertainty either as having major relevance (generalization error and parameter uncertainties), medium relevance (assignment error, transfer error and documentation error) or as having minor relevance (experts' evaluation error, procedural measurement error and approximation error). In the following, we will briefly introduce these different types of errors in this order of relevance.

Table 2.3: Sources of uncertainty in exposure assessment in the Wismut cohort adapted from Küchenhoff et al. (2018, p. 77).

Source of uncertainty	**Type**
Generalization error	Spatial Temporal
Assignment error	Spatial Temporal
Estimation error	Procedural error Documentation error Parameter uncertainties Experts' evaluation error Approximation error Transfer error

2.9.1 Generalization error

Generalization error occurs whenever a number of error-prone exposure measurements are averaged to determine the exposure estimate for a mining location. This type of error mainly occurred for radon gas and radon progeny concentration measurements, i.e.:

- Radon gas concentration measurements C_{Rn}
- Radon progeny concentration measurements C_{RDP}

As pointed out by Küchenhoff et al. (2018, p. 79), generalization error can arise through temporal, spatial and activity generalization errors and it is best described by a classical measurement error component. For years in which there was exposure monitoring, this classical measurement error component will merely be shared between all workers of the same object o as it affects all workers in this object in a similar way. For the years in which exposure assessment in the cohort was based on a retrospective estimation by experts (e.g. 1946 - 1954/55 for underground mining), on the other hand, the generalization errors arising in exposure assessment will not only be shared for all workers of the same objects o, but for all workers and all exposure years for which these exposure values were used: the classical measurement error component will therefore be shared both between and within workers. The same situation may arise when the exposure values assessed in a specific year are extrapolated to subsequent years. This extrapolation frequently occurred when exposure estimates were transferred to lead or to follow up times during the development of JEM 2.

2.9.2 Parameter uncertainties

The second source of error which was identified by Küchenhoff et al. (2018, p. 114) as having major relevance in the Wismut cohort are parameter uncertainties. The following parameters were involved in exposure assessment in the cohort:

1. Mined vein area $C(t, o)$
2. Evaluation area $A(t, o)$
3. Depth $d(t, o)$
4. Total shaft output $F(t, o)$
5. Density of the bedrock $h(o)$
6. Relative uranium recovery rate $r(t, o)$
7. Amount of uranium recovery $R(t, o)$
8. Void volume $V(t, o)$
9. Proportion of exposure from old mining in comparison to object 003 Schneeberg $b(o)$
10. Ventilation correction factor $c(o)$
11. Evaluation factors $e(p_t, o)$ and $e_2(p_t, o)$
12. Activity weighting factor $f(o, j)$
13. Equilibrium factor $g(p_t, o)$
14. Working time factor $w(p_t)$
15. Processing stage specific object weighting factor $z(o, j)$
16. Proportion of mined vein area from previous years p

Küchenhoff et al. (2018, p. 84) categorize these parameters either as parameters based on measurements (parameters 1 - 8 in the list given above) or as parameters based on aggregated evaluation of experts (parameters 9 - 15 in the list given above).
While Küchenhoff et al. (2018, p. 84) describe the parameter uncertainties arising in the first group of parameters, which are based on measurements, as classical measurement error and the parameter uncertainties arising in the second group of parameters arising through the evaluation of experts as Berkson error, we propose here to consider both a classical and a Berkson component for both types

of parameter uncertainties. In this vein, we will consider that for each uncertain parameter p involved in exposure assessment in the Wismut cohort, we have a measurement process of the true parameter value π, which is prone to classical measurement error. Depending on the parameter and the situation, the classical error component in this measurement process can either be shared for all miners for whom this parameter intervened in exposure estimation, it can be shared between all miners in a given mining location o and for all years of exposure assessment for these miners, or it can be only shared for all miners in a mining location o for a given year.

1. $p = \pi \cdot U_{\pi,c}$
2. $p(o) = \pi(o) \cdot U_{\pi,c}(o)$
3. $p(t,o) = \pi(t,o) \cdot U_{\pi,c}(t,o)$

For the two first situations, there will be an additional component of Berkson error, which can be considered to be shared between workers in the same mining location o for a given year t:

$$\begin{aligned} \text{for 1.: } \pi'(t,o) &= \pi \cdot U_{\pi',B}(t,o) \\ \text{for 2.: } \pi'(t,o) &= \pi(o) \cdot U_{\pi',B}(t,o) \end{aligned}$$

2.9.3 Assignment error

Assignment error arises whenever exposure is not determined for each worker individually, but jointly for an entire group of workers. As there was no use of personnel dosimetry in the Wismut cohort, contrary to the French cohort of uranium miners (Allodji et al., 2012a), for instance, assignment error occurred for all workers and for all exposure assessment periods in the Wismut cohort. Assignment error describes how the individual exposure that miner i received in year t, activity j and object o, $X_i(t,o,j)$, deviates from the observed exposure to radon progeny for year t, object o, and activity j $Z(t,o,j)$. Assignment error is best described by an additive Berkson error component because the true individual exposure values for the workers in a given group vary around the observed exposure value which was determined for this group and the resulting measurement error is independent of observed exposure. Küchenhoff et al. (2018, pp. 80–81) distinguish spatial, temporal, activity and working time assignment error in the Wismut cohort.

2.9.4 Transfer error

Transfer error occurs whenever missing exposure values were imputed based on exposure values either from another shaft or object or from another year in the same object in the absence of measurements, for instance during lead and follow-up times. Transfer errors can be seen as a special type of parameter uncertainty where the parameter to be estimated was a multiplicative constant which typically takes values between 0 and 2. Similar to other uncertain parameters, we will distinguish a Berkson and classical component for transfer errors and we will focus our attention on the following three transfer errors:

- transfer error for the evaluation factor t_e
- transfer error for the exposure of development and Saxonian exploration objects t_E
- transfer error for filled gaps in the exposure assessment according to the JEM (Lehmann et al., 1998; Lehmann, 2004) t_{JEM}

2.9.5 Documentation error

As pointed out by Küchenhoff et al. (2018, p. 83), documentation error may occur both in the documentation of measurements and in the documentation of individual occupational histories. In this sense, the following quantities could all be subject to documentation error:

- Radon gas concentration measurements C_{Rn}

- Radon progeny concentration measurements C_{RDP}
- Exposure histories for individual miners resulting in $l(i, t, o, j)$.

While unintentional documentation error is best described by a classical measurement error component, intentional documentation error may lead to systematic measurement errors.

2.9.6 Experts' evaluation error

Küchenhoff et al. (2018, pp. 85 - 86) distinguish three situations in which experts' evaluation error may occur in exposure assessment in the Wismut cohort: Either when experts modified some measured exposure values based on their knowledge concerning the ventilation conditions in certain mining locations, when exposure values were derived as proportional exposure values from another shaft or object or when exposure estimates were exclusively determined based on expert knowledge in the absence of any exposure measurements. Here, we mainly assumed experts' evaluation error for the latter case in which exposure estimates were determined through expert knowledge and it therefore applies for radon progeny exposure E in the measurement model M4 assumed for surface areas affiliated to mining objects and for exploration objects in Thuringia. While Küchenhoff et al. (2018, p. 86) describe experts' evaluation error through a Berkson error model, we will here consider this type of error as classical measurement error. Indeed, when exposure estimates are estimated based on expert knowledge, we are faced with a situation in which there is a true exposure value which is unknown and this unknown exposure value has to be estimated by experts. In this sense, the experts can be seen as a very imprecise measurement device and errors in the estimation process will affect the value of the observed exposure while being independent of the true exposure value, thereby satisfying the properties of a classical measurement error.

2.9.7 Procedural measurement error

Procedural measurement error occurs when exposure values are determined through measurement devices. In the Wismut cohort, we would expect procedural measurement error in particular for:

- Mean radon gas concentration measurements C_{Rn}
- Mean radon progeny concentration measurements C_{RDP}.

However, these quantities were in general the result of the averaging of concentration measurements. Therefore, the resulting measurement error is adequately described through generalization error and modelling both generalization error and procedural measurement error would be redundant. As described by Küchenhoff et al. (2018, p. 81), procedural measurement error is best described as classical measurement error and four different types of procedural measurement errors can be distinguished: methodological errors, calibration errors, statistical errors and human errors.

2.9.8 Approximation error

Approximation error may either arise through the use of rounding during the exposure estimation process or through the use of estimation equations (Küchenhoff et al., 2018, p. 87). Contrary to Küchenhoff et al. (2018, p. 87), we will consider approximation due to rounding as Berkson error as the error arising in this rounding process is independent of the observed exposure value rather than being independent of true exposure. Indeed, for a given observed exposure value Z, it is impossible to predict the precise value of the true exposure X. For an observed value Z of 5, for instance, it is impossible to predict the value of true exposure X or the extent of measurement error. Given a true exposure value X, on the other hand, both the observed value and the extent of the rounding error are known perfectly. For instance, for a given true exposure value of 4.7349238, rounding to an entire value will lead to an observed exposure of 5 and an error of -0.2650762. For the situation where rounding is performed to entire values, approximation error due to rounding can therefore be described by an additive Berkson error model of the form

$$X = Z + U$$

with U independent of Z.

Concerning the use of estimation equations, it is more difficult to characterize the resulting uncertainty as either Berkson or classical measurement error. This uncertainty arises from the fact that models were used to reconstruct exposure estimates in the absence of direct measurements of the quantity of interest. Models are necessarily simplifications and approximations of more complex phenomena. The resulting uncertainty can therefore be conceived as model uncertainty. Küchenhoff et al. (2018) describe the error resulting from the use of estimation equations as Berkson error.

3 Quantification of the magnitude of measurement error

In this chapter we will quantify the different measurement errors arising in the exposure assessment in the Wismut cohort. Firstly, we identify relevant literature sources and then describe the concept for the quantification of exposure uncertainty. In the third section we limit the effort and then quantify the different sources of uncertainty in the fourth section. Finally, the results of the quantification and the relevance of the different sources of uncertainty are discussed.

3.1 Relevant literature sources for the quantification of exposure uncertainty in the Wismut cohort

- Dosimetric reports of mining object 009 Aue for several years between 1961 and 1988 can be used for the quantification. The depicted information in the reports for 009 Aue varies depending on the year. Generally, the following three types of dosimetric reports can be differentiated.
  - For some years of object 009 Aue the dosimetric reports show the mean radon gas concentration for each shaft and level for both work place types mining and development. One example would be the year 1961.
  - The dosimetric reports of other years show the mean and the quantiles of radon gas and radon progeny concentration for each shaft and level for mining and development and state the shaft and level specific number of measurements. This is the case e.g. for year 1968 of object 009 Aue.
  - Some dosimetric reports only show the work place type specific mean radon progeny concentration and quantiles. This information is for example available for the years 1983 or 1988.
- Eigenwillig (2011) evaluates the radiation exposure assessment in the Wismut. Besides, the radon gas concentration measurements for object 009 Aue for 1955, 1956 and the first half of 1957 are reported.
- Lehmann et al. (1998) describe the exposure assessment in the Wismut cohort and depict the resulting JEM 1. This literature source also provides detailed information on radon gas and radon progeny concentration measurements and the assumed parameter values for the different years, objects and activities.
- Lehmann (2004) extends the object specific JEM (JEM 1) in Lehmann et al. (1998) to a shaft specific JEM (JEM 2). While Lehmann (2004) is not strictly relevant for the quantification of measurement error in the Wismut cohort, it is very important for the characterization of measurement error in this cohort and therefore listed here.
- Wismut GmbH (1999) is a detailed chronic of the Wismut company and reports among others changes in the object structure, the working conditions and the measurement of the radon gas and radon progeny concentration. The reported information about the mining losses over the years can be used for the quantification of the uncertainty in the proportion of mined vein area from previous years.
- Eigenwillig and Ettenhuber (2000) give among other things an estimate for the procedural measurement error arising for the measurement devices commonly used in uranium mining.
- Richter (1994) reports disruptions in the ventilation system as a frequent failure in underground mines that has not been correctly accounted for in the reported concentration values.
- Zettwoog (1981) describes results of the first experiments performed on the use of individual dosimetry in uranium mines in France and provides information on the comparison of ambient measurements through radon grab sampling and personal dosimetry in these mines.
- Allodji et al. (2012a) assess the extent of exposure uncertainty in the French cohort of uranium miners. They distinguish three distinct exposure assessment periods in this cohort, which are

characterized by a retrospective assessment of exposure values through a group of experts (1946-1955), by ambient measurements through scintillation flasks (1956-1982) and by personal dosimetry (1983-1999). The authors identify six sources of uncertainty, they assess the magnitude of these sources of uncertainty for the second and the third exposure assessment period and derive a combined measurement error variance for each period using a root sum square method. While the different sources of uncertainty arising in the error periods were of Berkson and classical and of shared and unshared nature, Allodji et al. (2012a) decide to consider all error occurring before 1983 in the French cohort of uranium miners as unshared Berkson error and all error occurring in the exposure assessment starting in 1983 as unshared classical measurement error. The authors consider all exposure uncertainty arising in the French cohort of uranium miners to be of a multiplicative nature and describe all exposure measurement errors by a lognormal distribution.

- Allodji et al. (2012b) study the impact of the exposure measurement errors characterised and quantified by Allodji et al. (2012a) based on a simulation study conducted on the French cohort of uranium miners. In the absence of an extensive assessment of the extent of exposure measurement error for the first exposure assessment period, the authors decide based on advice from experts specialized in radiation monitoring from the Algade company that the extent of exposure uncertainty in this period can be described by a lognormal distribution with a standard deviation of log-transformed values of 0.936.
- Schiager et al. (1981) characterize and quantify the extent of different sources of uncertainty arising in exposure monitoring of uranium miners in the United States.
- Lubin et al. (1995b) assume a multiplicative and lognormal error model when describing measurement error in radon exposure in studies on residential radon.
- Heid (2002) provides a justification for a multiplicative error model by arguing that many factors modifying radon concentration measurements affect a proportion rather than the absolute number of radon atoms in a given environment. Further evidence for a multiplicative error structure in radon measurements is provided in Heid et al. (2004) where the authors analyze data on the within- and between-laboratory-variability in an intercomparison study of laboratories in different European countries measuring radon gas concentrations in five houses and from three data sets with replicate radon concentration measurements based on alpha track detectors in German dwellings.

3.2 Concept for the quantification of exposure uncertainty in the Wismut cohort

In the following, we will describe the general concept for the quantification of exposure uncertainty in the Wismut cohort. Since there were neither a validation, calibration or replication study to obtain ancillary information to assess the nature and magnitude of measurement error, the quantification of exposure uncertainty in the Wismut cohort is not straightforward. It is particularly difficult to derive a concept for the quantification of measurement error for parameter uncertainties, because these are sometimes not based on any measurements, but rather on expert judgement. As a consequence, it was not possible to derive a convincing concept for some of the uncertain quantities (in particular for the parameter $b(o)$ quantifying the proportion of exposure from old mining in comparison to object 003 Schneeberg intervening in measurement model M1a and the classical error component for the activity weighting factor). In cases in which we describe several alternative quantification strategies, we chose to implement the strategy that appeared to be the most feasible and/or the most convincing strategy.

3.2.1 Generalization error

As described in Küchenhoff et al. (2018), generalization error arises through the usage of annual exposure estimation based on C_{Rn} or C_{RDP} instead of the actual annual shaft or level specific exposure. The standard deviation of this generalization error can be quantified with the help of

dosimetric reports of the Wismut. In Section 6.2.1 of Küchenhoff et al. (2018) the size of the generalization error is estimated by the variability of the annual exposure to radon progeny of a hewer in object o. On pages 98 and 99 in Küchenhoff et al. (2018) this is further approximated using the weighted sum of the variability of the mean radon gas or radon progeny concentration measurements at mining work places and at development work places. The only difference between radon gas and radon progeny concentration measurements is that for radon progeny concentration instead of the equilibrium factor $g(t, o)$ the correction factor for disruptions in the ventilation system $c(o)$ has to be used (Lehmann et al., 1998). We simplify the calculation of Küchenhoff et al. (2018) for the approximate size of the generalization error by not differentiating between mining or development work places for the mean radon gas or radon progeny concentration measurements. Also, since we are only interested in quantifying the generalization error for the radon gas or radon progeny concentration measurements, contrary to Küchenhoff et al. (2018) we omit the factors 12, $g(t, o)$, $c(o)$ and $w(t, o)$ in the standard deviation calculation.

The described calculation for the approximate size of the generalization error can be done for all years and objects for which the necessary information through e.g. dosimetric reports exist. As already mentioned in Section *3.1*, only dosimetric reports for some years of object 009 Aue are available.

3.2.2 Parameter uncertainties

General considerations

As described in Section *2.9.2*, we assume for every uncertain parameter a Berkson and a classical measurement error component. The classical measurement error describes the precision of the measurement process of this uncertain parameter and the Berkson component describes the variability between several years and/or objects for which the same parameter value was assumed.

A general strategy to quantify the Berkson component of an uncertain parameter is to quantify the variability of parameter values between years and between objects for the situation in which there were measurements of this parameter. As shown by Küchenhoff et al. (2018, p. 104), this strategy can for instance be used to assess the temporal variability in the equilibrium factor in object 009 Aue.

The classical measurement error component of an uncertain parameter is more difficult to assess. The estimation of the most influential parameters in the Wismut cohort, i.e. the equilibrium factor, the ventilation correction factor, the working time factor and the activity weighting factor were at least partially based on expert knowledge and it is very difficult to quantify the precision of this expert knowledge because we cannot rely on a calibration sample in which the true parameter values and the parameter values estimated by the experts could be compared. In this situation, a general strategy to quantify the uncertainty with which these parameter estimates can be determined consists in expert prior elicitation (Kadane and Wolfson, 1998; O'Hagan et al., 2006; Oakley et al., 2010; Fischer et al., 2017; Wolfson and Bousquet, 2016). In an expert prior elicitation task, an expert is not only asked to provide an estimate of an unknown quantity but also a measure of his uncertainty on this estimate. In the French cohort of uranium miners, Hoffmann (2017) performed an expert prior elicitation task with three experts on the working conditions in French uranium mines to obtain information on the time workers spent in different levels of physical activity. Based on the elicited information, the authors derived an informative prior distribution on average breathing rate in the French cohort of uranium miners to account for dose uncertainty when analyzing the association between absorbed lung dose and lung cancer mortality in this cohort. Similarly, if we could contact experts on the working conditions in the Wismut cohort, it would be possible to make an expert prior elicitation with them. Alternatively, we could also ask experts to estimate values of parameters for which measurements exist, but that these experts do not know. In this situation we can determine the precision in expert knowledge by comparing the true parameter values (which are only known to us) and the values provided by the experts. Finally, we can use the

information which is available concerning the imprecision that was acceptable to the experts who were involved in the estimation of exposure values for the Wismut cohort. Lehmann et al. (1998, pp. 123-124) provide such information concerning the evaluation factor.

Evaluation factor $\boldsymbol{e(o)}$

Evaluation factors occur both in M1a, M1b and in M6. As the evaluation factors in M1a and M1b are deterministic functions of other quantities, i.e. $\tau_e(t, o)$, $\mathcal{C}_{Rn}(t_0(o_0(o)), o_0(o))$ and $A(t_0(o_0(o)), o_0(o)$, we do not need to quantify the uncertainty on the evaluation factor in these models directly. However, since we have some information on the evaluation factor in these models whereas we lack information for the evaluation factor in M6, it is advisable to use this information for the quantification of uncertainty for the evaluation factor in M6.

In order to assess the Berkson component of the parameter uncertainty in the evaluation factor, we could calculate the ratio between the mean radon gas and radon progeny concentration, which was measured in the second and the third exposure assessment period, and the mined area in these years. If information on the mined area in these years was available, we could calculate the variability in these ratios for different years and objects for which the same evaluation factor was assumed in the exposure assessment in the first exposure assessment period. Unfortunately, there are no longer any records of the mined area, which is why this approach to quantifying the Berkson error component of the evaluation factor does not work.

To assess the classical error component, we can use information provided by Lehmann et al. (1998, pp. 123-124) concerning the imprecision in the determination of the evaluation factor which was acceptable to the experts.

Equilibrium factor $\boldsymbol{g(p_t, o)}$

The values of the equilibrium factor which were assumed for different exposure years and the different objects are given in Table 10 in Küchenhoff et al. (2018, p. 45). When quantifying the uncertainty in the equilibrium factor we can distinguish a Berkson and a classical measurement error component.

In order to quantify the Berkson component of the equilibrium factor, we can assess the variability of the equilibrium factor for situations in which this factor can be determined through measurements. In order to assess the temporal variability in the equilibrium factor, Küchenhoff et al. (2018, p. 104) estimate equilibrium factors for the years between 1966 and 1981 where radon gas and radon progeny concentrations were measured simultaneously in the object 009 Aue. Following this idea, it is possible to assess the variability between several objects and years for which the same value was assumed for the equilibrium factor (see Table 10 in Küchenhoff et al. (2018, p. 45)) by estimating the equilibrium factor for several objects and years based on the estimates for radon gas and radon progeny concentrations which are given in Lehmann et al. (1998, pp. 436–463). Besides, information on the variability in a given year and object is provided in the dosimetric report for the object 009 Aue for the year 1962: This report states that for shaft 038 with measurements of $g(t, o)$ between 11% and 55% an equilibrium factor of 25% was assumed. Similarly, for shaft 366 and 371 with measurements of $g(t, o)$ between 10% and 100% an equilibrium factor of 50% was assumed.

The quantification of the classical component of the measurement error of the estimation of the equilibrium factor is more intricate than the quantification of its spatial and temporal variability. One possible strategy to assess the precision of the estimation of the equilibrium factor would be to use the dosimetric reports of years with shaft and level specific radon gas and radon progeny concentration measurements. For this possibility to quantify the equilibrium factor, only the dosimetric reports with radon gas and radon progeny concentration measurements for the years until 1973 and from 1987 to 1990 should be used. In the years between 1974 and 1986 there was a problem with the correct handling of the radon progeny concentration measurement devices, which

lead to incorrect concentration measurement values (Wismut GmbH, 1999, Chapter 1.8.2, p. 23). Thus, the calculation of the equilibrium factor for these years would be biased. The year 1968 of the Saxonian object 009 Aue is one example where the dosimetric report contains the radon gas and as well the radon progeny concentration.
With radon gas and radon progeny measurements in the same shaft and at the same level it is possible to calculate the shaft and level specific equilibrium factors by

$$\text{equilibrium factor} = \frac{\text{radon progeny concentration in WLM}}{\text{radon gas concentration in Eman}}.$$

The standard deviation of the shaft and level specific equilibrium factors calculated respectively, can be used as an approximation of the precision of the estimation of the equilibrium factor. The described approach quantifies an additive error component but only needs the following slight adaption to quantify a multiplicative error component: Instead of simply taking the standard deviation of the shaft and level specific equilibrium factors, the standard deviation of the log-transformed values is used to approximate the precision of the estimation of the equilibrium factor for a multiplicative error.

Another possibility would be to base the quantification of the precision of the estimation of the equilibrium factor on values described in the literature for other cohorts of uranium miners. For the French cohort of uranium miners, Allodji et al. (2012a) describe the uncertainty in the approximation of the equilibrium factor by a lognormal distribution with a standard deviation of log-transformed values of 0.294 for the period between 1956-1977 and of 0.118 for the period 1978-1982.

Working time factor $w(p_t)$

While there was information on the number of annual working hours in the literature for the exposure years between 1966 and 1990, this number had to be determined through interviews with experts for the exposure years between 1946 and 1965 (Küchenhoff et al., 2018, p. 44). Therefore we could theoretically assume that the classical measurement error component arising through the imprecision of parameter estimation only occurred in the exposure years between 1946 and 1965, while the Berkson component arising from the variability in the values for different years and different objects around a common parameter value was present for all exposure years. However, for the sake of simplicity and since it is unlikely to lead to noticeable changes in the estimation of the effect of radon exposure on the development of lung cancer, we assume a classical and a Berkson measurement error component on the working time factor for the entire time.
Concerning the Berkson component, a theoretical possibility would be to rely on information in the French cohort of uranium miners. In this cohort, there were detailed transcripts of the number of hours the workers worked in the mines and we could assess the variability in these working hours to extrapolate this information to the Berkson component in the working time factor in the Wismut cohort. However, quantifying the Berkson component for the working time factor in this way would require access to internal documents on the French cohort of uranium miners and the effort of processing these additional documents was not in the scope of this project. In the absence of more information in the Wismut cohort concerning the classical error component, we can use information provided by Schiager et al. (1981) and Allodji et al. (2012a) for uranium mines in the United States and in France.

Activity weighting factor $f(o, j)$

In the French and in the Czech cohort of uranium miners, there were exposure periods during which individual exposure measurements were obtained via personal dosimetry. If these individual measurements could be obtained, it would be possible to quantify the Berkson component in the activity weighting factor by assessing the ratio between the average exposure values for a given activity j and the mean exposure of hewers for several years in the French and in the Czech cohort of

uranium miners. In the absence of information from these cohorts, we cannot use data from the Wismut cohort, because there were no periods of individual exposure measurements in the Wismut cohort. However, we could try to provide a range of plausible values for the activity weighting factor by considering the fact that this parameter can only take values between zero and one. As this last strategy would give us only very imprecise estimates on the magnitude of error we can assume when correcting for measurement error in the Wismut cohort, it seems more reasonable to make the assumption that the magnitude of uncertainty in the activity weighting factor is comparable to another parameter in the Wismut cohort for which we can quantify the magnitude of Berkson error.

Ventilation correction factor $c(o)$

The ventilation correction factor $c(o)$ corrects the exposure values to account for disruptions and breakdowns of the ventilation system. The more disruptions there are, the higher becomes the radon exposure. If the measured radon progeny concentration values were too high, the Wismut company issued warnings, that sometimes lead to closure of the shaft. Some dosimetric reports of object 009 Aue, for instance for 1968, give the number of warnings and closures of shafts. One possibility to quantify the Berkson component of the ventilation correction factor would be to use the variability of the warnings and closures in one object and year over the different levels and shafts. Note that the range of values of the ventilation correction factor and the range of values of the number of warnings and closures are quite different. However, this difference in the ranges of values is not a problem when quantifying a multiplicative error component. Another possibility would be to base the quantification on Richter (1994) where the reported exposure values are corrected by factors of 1.3 to 1.7 to account for breakdowns in the ventilation system.

Proportion of mined vein area from previous years p

Lehmann et al. (1998) describe that the proportion p of the mined vein area $C(s, o)$ from previous years ($s < t$) intervening in the calculation of the evaluation area $A(t, o)$ of year t for objects in Saxony was assumed to be 0.2 since the average of the mining losses was about 20%. However, it can be assumed that the factor of 0.2 is prone to error since there was a variability over the years. In Wismut GmbH (1999, Chapter 2.2.2.1, p. 15) an overview of the proportions of losses over the years can be found. They vary between 0.62 in 1949 and 0.08 at the end of the uranium mining activity in the Wismut. Using these proportions of mining losses, it is possible to approximate the classical error component of the proportion p of the mined vein area $C(s, o)$ from previous years ($s < t$) by the variability of the log-transformed values depicted in Wismut GmbH (1999, Chapter 2.2.2.1, p. 15).

3.2.3 Assignment error

Since the JEMs for the Wismut cohort use object (JEM 1; Lehmann et al. (1998)) or shaft specific (JEM 2; Lehmann (2004)) exposure values instead of the individual exposure measurements for each miner, an assignment error occurs (Küchenhoff et al., 2018).
As there were no individual exposure measurements in the Wismut cohort, it is impossible to directly quantify the assignment error. Due to this absence of individual exposure values, Küchenhoff et al. (2018, p. 102) propose to use the dosimetric reports of the Wismut as source for the quantification. They argue that the magnitude of spatial assignment error can be assessed by quantifying the variability of shaft/object specific exposure estimates since information on single measurements of radon gas and radon progeny concentration are missing in the available dosimetric reports. Küchenhoff et al. (2018) average the within-group variability of the site or object specific exposure values, whereby they assume equal numbers of measurements within each group whenever a more detailed information is not available (see equation on page 103 of Küchenhoff et al. (2018)).

A second strategy to obtain an estimate of the magnitude of assignment error in the Wismut cohort is to determine the variability of individual exposure values for a given year and a given shaft/object in the French and in the Czech cohort of uranium miners for the years in which exposure to radon

progeny was assessed through individual dosimetry by personal alpha dosimeters, i.e. starting in 1983 in the French cohort (Allodji et al., 2012a) and starting in 2000 in the last operating mine in the Czech republic (Marušiaková et al., 2011).

A third strategy would be to return to the mining locations and to simultaneously measure with area dosimetry and personal dosimetry to quantify the magnitude of assignment error. Such measurements were performed by Zettwoog (1981) in French uranium mines, but he merely reported the overestimation of exposures through ambient measurements and not the variability of results and the data he collected are not available. Eigenwillig and Ettenhuber (2000) also mention data collected during 2 years of comparisons between ambient measurements and personnel dosimetry and state that these comparisons provided evidence that exposure estimates provided by ambient measurements exceeded the exposure estimates based on personal dosimetry by 10%. Details on the space, time and setting of this comparison or information on where the data of these comparisons could be found were not available.

3.2.4 Transfer error

To quantify the classical error component in the transfer error for the evaluation factor τ_e, we can assess to what extent the exposure values in an object were similar to the exposure values of its reference object for years in which these exposure values were measured. However, the problem in this quantification strategy would be that we should not only calculate the variability in the exposure values, but we should also account for the evaluation area $A(t, o)$ for these latter years and we are lacking this information.

To quantify the Berkson component in the transfer error for the exposure of development and exploration objects τ_E, we could try to use information which could be available in the French and in the Czech cohort of uranium miners. In these cohorts, there could be exposure years where there was a systematic and prospective exposure assessment through ambient measurements or personal dosimetry in exploration and development objects. Based on this information, we could assess the variability in the ratio between these exploration and development objects and the corresponding mining objects and extrapolate these values to the Wismut cohort.

Concerning the transfer error for filled gaps in the exposure assessment according to the JEM, it is possible to quantify both the Berkson and the classical error component by comparing the measured concentration values for the years in which there were measurements (Lehmann et al., 1998, p. 437). For the classical measurement error component, during these years, we can assess what error would have been made if we had replaced any of the available values by a measurement that was performed in any of the other years, i.e. for a given year t, we assume that the actual measured exposure value is the true exposure value X and consider the values in the years before and after this year t as potentially observed exposure value to assess the magnitude of classical measurement error for the transfer error for filled gaps in the exposure assessment.
To assess the Berkson component which describes the variability in the transfer errors for several years for which the same value was assumed in the filling of the gaps, we can also use the measured exposure values for the years in which there were observed measurements. In this vein, we can quantify the variability around the mean exposure value for an object in which these exposure values were actually measured for several years. More precisely, we calculate the mean value of the yearly exposure values for a given object and determine the variability of the yearly measured exposures around this mean to mimic the Berkson error that would arise if we assumed a constant exposure value for several subsequent years of exposure in an object. Quantifying the transfer error when exposure values were transferred from one object to another object is not possible since we are lacking measurements for the latter objects. We could in theory just make the assumption that the measurements in the reference objects were as different from the measurements that would hypothetically been obtained in the objects these values were transferred to, but then we have to

answer the question why these values in particular were used to replace the unmeasured values. As a consequence, it seems to make more sense to use the standard deviations quantified for the temporal transfer error also for the latter situation.

3.2.5 Documentation error

In the absence of precise information on the magnitude of documentation error in the Wismut cohort, we can rely on information provided by Schiager et al. (1981) and Allodji et al. (2012a) on the magnitude of documentation error in uranium mines in the US and in France.

3.2.6 Experts' evaluation error

Similar to documentation error, we can rely on information provided by Allodji et al. (2012a) on the magnitude of experts' evaluation error in uranium mines in France.

3.2.7 Procedural measurement error

In order to quantify procedural measurement error in the Wismut cohort, the best option would be to return in the old mining locations and to perform again measurements with the same type of measurement devices that were used to determine the exposure values of the cohort. These measurements could then be compared with a gold standard, i.e. with a measurement device which is very precise and presents only negligible measurement error to derive the variance of this classical measurement error. Alternatively, we could also compare several measurements with the same error-prone measurement device. This would allow us to assess the retest reliability of this device and to derive the corresponding measurement error variance.
In the absence of measurements which could inform us about the precision of the measurement devices in the Wismut cohort, it is possible to rely on external literature sources, which we described in Section *3.1*, namely Schiager et al. (1981), Eigenwillig and Ettenhuber (2000), Zettwoog (1981) and Allodji et al. (2012a) who assess the extent of measurement error in the United States and in the French cohort of uranium miners, respectively.

3.2.8 Approximation error

Approximation error due to rounding

In order to assess the extent of approximation error due to rounding, it is possible to compare the measured values in the archives of the Wismut cohort with the assumed values in the JEM. In general, the quantification of the magnitude of error arising through rounding is quite straightforward as it is merely a function of the number of digits to which rounding is performed. As described in Section *2.9.8*, for the situation where rounding is performed to entire values (which was not always the case in the Wismut cohort), approximation error due to rounding can be described by an additive Berkson error model of the form

$$X = Z + U$$

with U independent of Z.

Approximation error arising from the use of estimation equations

Concerning approximation error arising from the use of estimation equations, Küchenhoff et al. (2018, p. 87) characterize the resulting exposure measurement error, which essentially arises from the fact that it was necessary to rely on models to reconstruct exposure values, as Berkson and classical measurement error. However, as mentioned in Section *2.9.8*, it can be argued that this second type of approximation error is best understood as model uncertainty in exposure assessment. It is very likely that a group of experts would be able to think about several alternative models for the assessment of exposure values. In this sense, a way to quantify this type of approximation error would be to ask experts (either the original experts who were involved in the assessment of radon

exposure in the Wismut cohort or new experts, for instance dosimetrists who specialize in radon dosimetry in uranium mines) for alternative modelling strategies which could have been possible to reconstruct these individual exposure values and for estimations of the probability of each model to be the true model for exposure assessment. Once these alternative models are obtained, it would be possible to account for this type of approximation error for instance through Bayesian Model Averaging (Hoeting et al., 1999; Volinsky et al., 1997).

3.3 Limiting the effort of exposure quantification in the Wismut cohort

As it would be infeasible to quantify and account for all sources of uncertainty identified by Küchenhoff et al. (2018) in the scope of this project, it is necessary to make a number of choices to limit the efforts in the quantification process while simultaneously accounting for the sources of uncertainty that have an impact on risk estimation in the cohort.
We made the following choices to limit the effort of exposure quantification in the Wismut cohort:

- While it would have been possible to quantify each type of uncertainty as additive measurement error component and as multiplicative error component, we decided to focus on the quantification of multiplicative errors based on a broad consensus in the literature that measurement error in radiation exposure in general, and in radon exposure in particular, is best described by a multiplicative error component following a lognormal distribution (Lubin et al., 1995b; Stram et al., 1999; Heid, 2002; Heid et al., 2002; Heid et al., 2004; Heidenreich et al., 2004; Lubin et al., 2005; Advisory Group on Ionising Radiation AGIR, 2009; Heidenreich et al., 2012; Allodji et al., 2012a,b,c). We also make this assumption for parameter uncertainty as a multiplicative error structure seems often reasonable and convenient when the true parameter can only take positive values. Ideally, this assumption of multiplicative errors should be complemented by extensive sensitivity analyses in which an additive error structure is assumed, but it was not in the scope of this project to perform these additional sensitivity analyses.
- Similarly, even if with additional assumptions it would have been possible for some error components to assume different error variances for at least some of the different years and different objects in a given measurement model, in order to limit the effort of exposure quantification, we only determined one magnitude of error for each measurement model for each quantity and extrapolate this value for all years and objects.
- As pointed out by Küchenhoff et al. (2018, p. 114) and confirmed in a simulation study on the impact of the different sources of uncertainty on a fictive data set of the Wismut cohort (results not shown), approximation error due to rounding only has a negligible effect on exposure estimates and on risk estimation in this cohort. Although it would be relatively easy to quantify this source of uncertainty (as described in Section *2.9.8*, the magnitude of approximation error merely depends on the number of digits to which rounding is performed), we will not consider this type of error in detail in the following.
- When it comes to the approximation error due to estimation equations, it seems very difficult to identify experts on the exposure conditions in the Wismut cohort, so it is difficult to determine alternative models to use in the assessment of exposure values. Moreover, it is not in the main scope of this project to account for model uncertainty. Therefore, it appears to be difficult from a methodological perspective to extent the approach that will be used to account for measurement error to model uncertainty arising from the fact that estimation equations were used to reconstruct exposure values and it appears to be more judicious to focus all efforts on the development of a suitable method to account for the complex structure of measurement error arising in this cohort. Thus, we will not consider the approximation error due to estimation equations in the quantification.
- In accordance with Küchenhoff et al. (2018, p. 114), we believe, that documentation error is only of minor importance in the Wismut cohort. Additionally, it would be very difficult to account for

documentation error in working histories as these working histories are not easily accessible in the Wismut cohort. We therefore did not account for documentation error in the following.

- Due to the lack of dosimetric reports for mining objects other than Aue, we will only perform the quantification of generalization error and assignment error for this object and extrapolate the magnitude of these two error components to all other mining locations. Additionally, the quantification of these two types of error is only possible for certain years, which means that we will also extrapolate the magnitude of these two error components over time for the years for which sufficient information for a quantification is not available.
- We agree with Küchenhoff et al. (2018, p. 114) that procedural measurement error only rarely occurred in the exposure assessment in the Wismut cohort and that it is thus not relevant to account for this source of classical measurement error. As mentioned in Section *2.9.7*, most measurements in the Wismut cohort were the result of the averaging of measured concentrations and the resulting measurement error is therefore better described through generalization error. As a result, we did not account for the uncertainty in the procedural error when accounting for measurement error in the Wismut cohort.
- As it seemed very difficult and time-consuming to identify experts on the working conditions in the Wismut cohort, we decided not to elicit prior distributions based on expert knowledge to quantify parameter uncertainty, even though this seemed the most adequate option to quantify the classical measurement error component arising in parameter uncertainties. One parameter for which it was possible to quantify this classical measurement error component arising from the imprecision in the expert estimation process was the evaluation factor in the first exposure assessment period for underground mining objects. We will assume that the precision of the estimation of parameter values is comparable for all uncertain parameters and extrapolate from the magnitude of the classical error component as quantified in Section 3.4.2 to the classical error component for all other uncertain parameters for which additional data that could inform us on the magnitude of the classical measurement error component is absent.
- In this work, we identified a number of different transfer factors. As we are lacking information to quantify the errors for all transfer factors except the transfer factor which occurred because years with missing exposure values in the JEM were filled by assuming that the exposure in these years was the same as in other years for which there were exposure values, we will quantify the error for this specific transfer factor t_{JEM} and extrapolate its magnitude to all other transfer factors.

3.4 Quantification of exposure uncertainty in the Wismut cohort

3.4.1 Generalization error

As described in Section *3.2*, following Küchenhoff et al. (2018), the generalization error for years with suitable dosimetric reports can be quantified by the variability of the mean radon gas or radon progeny concentration measurements. For example, the dosimetric reports from object 009 Aue for the years 1961 and 1968 and Table 9 from Eigenwillig (2011, p. 112) for the year 1955 were suitable. We chose these years because they were the first years for the three different exposure assessment periods for which the necessary information was available. We use the value for 1961 for all objects and years in the second exposure assessment period for underground mining objects and for open pit mining objects. We use the value for 1968 for all objects and years in the third exposure assessment period for underground mining objects.
For the year 1961 of object 009 Aue the size of the standard deviation of the additive generalization error occurring for the mean radon gas concentration $C_{Rn}(1961{,}009)$ is calculated as

$$\sqrt{\widehat{Var}(U_{\mathcal{C},c}(1961{,}009))} \approx 0.59.$$

With the dosimetric report for the year 1968 of object Aue, theoretically the calculation of the size of the generalization error would be possible in the exact same way as for the year 1961. However, since in 1968 object 009 Aue belongs to the third exposure assessment period with radon progeny concentration measurements, the mean radon progeny concentration measurements in 10^3 MeV/l first have to be divided by 130 to obtain the unit Eman. Thus, the size of the standard deviation of the additive generalization error for $C_{RDP}(1968{,}009)$ can be approximated as

$$\sqrt{\widehat{Var}(U_{\mathcal{C},c}(1968{,}009))} \approx 0.03.$$

The dosimetric report of object 009 Aue in 1955 can be used to quantify the generalization error occurring in the mean radon gas concentration $C_{Rn}(1955{,}009)$ after first dividing the shaft specific mean concentration measurements of radon gas by 3.7 in order to transform the unit from kBq/m^3 to Eman. The standard deviation of the additive generalization error for object Aue in 1955 can be approximated as

$$\sqrt{\widehat{Var}(U_{\mathcal{C},c}(1955{,}009))} \approx 5.29.$$

The generalization error that was quantified so far, is the generalization error for mining activity. Another type of generalization error is the generalization error for the basic exposure from old mining $C_{Rn}(1937/38{,}003)$ for objects in the first exposure assessment period in Saxony. The basic exposure from old mining is based on the measurements in 1937/38 of object 003 Schneeberg. Thus, the size of the generalization error for basic exposure from old mining can be quantified by using the radon gas concentrations from Table 4.1.2.1 on page 433 of Lehmann et al. (1998) without shaft 'Siebenschlehen' (Küchenhoff et al., 2018). As the unit of the radon gas concentrations in Table 4.1.2.1 is kBq/m^3, the mean concentration values have to be first divided by 3.7 in order to obtain the unit Eman. Then, the standard deviation of the additive generalization error for the basic exposure from old mining can be calculated as

$$\sqrt{\widehat{Var}(U_{\mathcal{C},c}(1937/38{,}003))} \approx 6.56.$$

3.4.2 Parameter uncertainties

Evaluation factor $e(o)$

As described in Section 3.2.2, we used the information provided by Lehmann et al. (1998, pp. 123-124) to assess the imprecision in the determination of the evaluation factor $e(o)$ in the exposure assessment in the first exposure assessment period for objects in Saxony which was acceptable to the experts: For the object BB Dittrichshütte and the object BB Hirschbach the uranium contents were 0.04% and 0.03%, respectively. The experts derived the same value for the evaluation factor for both mining locations by considering that both values were roughly half as large as the value for the uranium content for the object BB Schmirchau which was 0.083%. By doing so, they accepted a deviation of 33% for the object BB Hirschbach and we can conclude that this magnitude of error seemed normal to them. It can therefore be argued that the classical error component arising in the estimation of the evaluation factor can be described by a multiplicative error term following a lognormal distribution with a standard deviation of log-transformed values of 0.33 (Allodji et al., 2012a).

$$\sqrt{\widehat{Var}(U_{\epsilon,c})} \approx 0.33.$$

We will use this value of 0.33 that was quantified for the classical error component of the evaluation factor for the classical error components for all uncertain parameters that cannot be quantified more accurately.

Equilibrium factor $g(\boldsymbol{p_t}, \boldsymbol{o})$

The variability of the classical component of the equilibrium factor can be quantified with the help of parallel measuring of the shaft and level specific radon gas and radon progeny concentration in one specific year of a specific object. This information can be found in the dosimetric report for this year and object. Since only dosimetric reports are available for object 009 Aue, a quantification in this way can only be done for a few years for this object. Namely for the years for which both, radon gas and radon progeny measurements, are available. In the following the calculation is performed for the year 1968 of object 009 Aue. The measured radon gas concentration is given in 10^{-10}Ci/l and the measured radon progeny concentration in 10^3MeV/l. Since the equilibrium factor can be calculated by dividing the radon progeny exposure in WLM by the radon gas concentration in Eman, the radon progeny concentrations in the dosimetric report have to be transformed from 10^3MeV/l to WLM by dividing them by 130. The unit 10^{-10}Ci/l for radon gas measurements is basically the same as Eman. Thus, the equilibrium factors for object 009 Aue in 1968 can be calculated by

$$g(1968{,}009) = \frac{\text{radon progeny concentration in } 10^3\text{MeV/l}}{130 \cdot \text{radon gas concentration in } 10^{-10}\text{Ci/l}}.$$

For the calculation the columns of the dosimetric report showing the shaft and level specific mean concentration values for mining and development were used.

The multiplicative classical error component of the equilibrium factor for the year 1968 of object 009 Aue is calculated on the log-transformed level and shaft specific equilibrium factors as

$$\begin{aligned}\sqrt{\widehat{Var}(U_{\gamma,c})} &\approx \sqrt{\widehat{Var}(log(g(1968{,}009)))} \\ &\approx 0.23.\end{aligned}$$

As described in the concept in Section *3.2*, the Berkson error component of the equilibrium factor can be quantified with the variability of the calculated equilibrium factors between the objects and over the different years for which the same value was assumed for the equilibrium factor (see Table 10 in Küchenhoff et al. (2018, p. 45)). In Appendix 4.2.1 of Lehmann et al. (1998) the annual radon progeny concentrations in MeV/cm^3 and the annual radon gas concentrations in Eman are shown for some objects and years. To calculate the equilibrium factor, the radon progeny exposure in WLM has to be divided by the radon gas concentration in Eman. Thus, we need to transform the annual radon progeny concentrations in Appendix 4.2.1 of Lehmann et al. (1998) from MeV/cm^3 to WLM by dividing the values by 130. Hence, we calculate the equilibrium factor as

$$g(t, o) = \frac{\text{radon progeny concentration in MeV/cm}^3}{130 \cdot \text{radon gas concentration in Eman}}.$$

We use all objects and all years where both the mean radon gas concentration in Eman and the radon progeny concentration in MeV/cm^3 are available. As for all of these possible years, all objects except for object 009 Aue assume an equilibrium factor of 0.3 according to Table 10 in Küchenhoff et al. (2018, p. 45), we omit the concentration measurements of all years of object 009 Aue.

By log-transforming the calculated equilibrium factors of the remaining objects, we approximate the multiplicative Berkson error component of the equilibrium factor as

$$\begin{aligned}\sqrt{\widehat{Var}(U_{\gamma',B})} &\approx \sqrt{\widehat{Var}(log(g(t,o)))} \\ &\approx 0.69.\end{aligned}$$

We will use this value of 0.69 that was quantified for the Berkson error component of the equilibrium factor for the Berkson error components for all uncertain parameters that cannot be quantified more accurately.

Working time factor $w(p_t)$

Following Schiager et al. (1981), Allodji et al. (2012a) describe the classical error component in the uncertainty in the estimation of the working time factor through a lognormal distribution with a standard deviation of log-transformed values of 0.04 before the mechanization of work in the mines. After the mechanization of work in the mines, Allodji et al. (2012a) estimate an increase in the uncertainty in the estimation of the working time factor which they translate by a lognormal distribution with a standard deviation of log-transformed values of 0.08. As described in Section *3.2.2*, a classical measurement error component arising through the imprecision of parameter estimation theoretically only occurred in the exposure years between 1946 and 1965. As these exposure years are more comparable to the period before the mechanization of work in the mines in the French cohort of uranium miners than to the period after the mechanization of work in the mines, we will make the assumption that the classical error component arising in the estimation of the working time factor for the entire employment period of the Wismut cohort can be described by a lognormal distribution with a standard deviation of log-transformed values of 0.04:

$$\sqrt{\widehat{Var}(U_{\omega,c})} \approx 0.04.$$

Ventilation correction factor $c(o)$

The Berkson component of the error for the ventilation correction factor $c(o)$ can be quantified using the number of warnings, closures and direct closures in the dosimetric report of object 009 Aue in 1968. We quantify the standard deviation of the multiplicative error of the ventilation correction factor as

$$\begin{aligned}\sqrt{\widehat{Var}(U_{ç',B})} &\approx \sqrt{\widehat{Var}(log(n_{warnings})) + \widehat{Var}(log(n_{closure})) + \widehat{Var}(log(n_{direct}))} \\ &\approx 1.45.\end{aligned}$$

Proportion of mined vein area from previous years p

The classical error component of p can be quantified using the values of the mining losses over the years in Wismut GmbH (1999, Chapter 2.2.2.1, p. 15). The size of the standard deviation of the multiplicative classical error can be obtained by the variability of the log-transformed proportions of losses as

$$\sqrt{\widehat{Var}(U_{\pi,c})} \approx 0.79.$$

3.4.3 Assignment error

Following Küchenhoff et al. (2018), the magnitude of the additive assignment error can be calculated by averaging the within-group variability of the level and shaft specific concentration measurements. This is only possible for years for which dosimetric reports with the necessary information exist, which is the case for the year 1961. For object 009 Aue in 1961, Küchenhoff et al. (2018) estimate the size of the standard deviation of the assignment error arising on $E(t, o, j)$ as

$$\sqrt{\widehat{Var}(U_{E,B}(1961{,}009))} \approx 13.95.$$

We use this value for all years and objects in the Wismut cohort.

3.4.4 Transfer error

For the quantification of the classical measurement error component of the transfer error for filled gaps in the exposure assessment, we use the annual mean radon gas concentrations of object 009 Aue for all years depicted in Table 4.2.1.2 in Lehmann et al. (1998, p. 437) until 1965, since starting from 1966 radon progeny concentration measurements were used for exposure assessment. To

assess what error would have been made if instead of the measured concentration value in the respective year, the concentration value of the previous year or alternatively, the concentration value of the following year would have been used, we calculate the two ratios of 1) the concentration value from the previous year divided by the concentration value of the actual year and 2) the concentration value from the following year divided by the concentration value of the actual year. Then, we use the standard deviation of the log-transformed ratios in order to approximate the multiplicative classical error component of the transfer factor for filled gaps in the exposure assessment in the second exposure assessment period as

$$\sqrt{\widehat{Var}(U_{\tau_{JEM},c})} \approx 0.37.$$

The log-transformation in this calculation is necessary since we always parameterize the lognormal distribution as the mean and standard deviation of the underlying normal distribution of log-transformed values. Note that taking the exposure values of previous years and following years leads to some degree of redundancy in this calculation, because the ratio of two values a and b that were measured in subsequent years enters once as a/b and then as b/a in this calculation. As described in the quantification concept in Section *3.2*, we can approximate the Berkson error component of the transfer error for filled gaps in the exposure assessment, by the variability around the mean exposure value for an object in which the radon gas or radon progeny concentration values were actually measured separately for several years. We quantify the Berkson error component of this transfer error differently for measurement models M2, M3 and M4, by using either the radon gas concentrations (M2), the radon progeny concentrations (M3) or the radon exposures in WLM (M4) from object 009 Aue. Regardless of the used measurement model, we first calculate the mean values of the exposures for three consecutive years and then the ratios of the actual yearly exposures divided by these three-yearly mean exposure values. The multiplicative Berkson error component of the transfer factor for filled gaps in the exposure assessment can then be approximated by the standard deviation of the log-transformed ratios.

For the multiplicative Berkson transfer error in M2 all mean radon gas concentration values from Table 4.2.1.2 Lehmann et al. (1998, pp. 436–441) until 1965 are used to obtain the standard deviation of

$$\sqrt{\widehat{Var}(U_{\tau'_{JEM},B})} \approx 0.33.$$

For the quantification for M3 we use the radon progeny concentration values from Table 4.2.1.2 Lehmann et al. (1998, pp. 436–441) starting from 1966, which are depicted in the unit MeV/cm^3 and first need to be transformed to Eman by dividing them by 130 before the standard deviation of the multiplicative Berkson error for transfer factors in M3 can be approximated as

$$\sqrt{\widehat{Var}(U_{\tau'_{JEM},B})} \approx 0.13.$$

The multiplicative Berkson error for the transfer factor in M4 is quantified with the exposure values in WLM from Table 2.1.1.9/1 in Lehmann et al. (1998, pp. 220-221) for the years 1955-1957 and 1961-1988 as

$$\sqrt{\widehat{Var}(U_{\tau'_{JEM},B})} \approx 0.18.$$

As described in Section *3.2.4*, we use the standard deviations quantified for the temporal transfer error also for situations in which the exposure values measured in one object were transferred to another object since it would be very difficult to quantify the errors arising in this latter situation. Indeed, it is not clear which objects for which actual measurements exist would have been considered to be similar enough to transfer the measurements from one object to the other object

and just making the assumption that all objects are comparable might lead to an overestimation of this type of transfer error.

3.4.5 Experts' evaluation error

Based on the advice from experts specialized in radiation monitoring, Allodji et al. (2012b) describe the exposure measurement error arising from the estimation of exposure values by experts through a lognormal distribution with a standard deviation of log-transformed values of 0.936. In the absence of additional information on the precision of exposure estimates performed by experts in the exposure assessment in the Wismut cohort, we can assume that exposure assessment was similar in the two cohorts and adapt the values determined by Allodji et al. (2012b) for the measurement error $U_{\mathcal{E},c}$ arising for the uncertainty in $\mathcal{E}(p_t, o)$, $\mathcal{E}(o_0, j)$ and $\mathcal{E}^*(p(t, o), o_0, j)$.

3.5 Result of the quantification and relevance of the different sources of uncertainty

When quantifying and accounting for uncertainty in exposure assessment in the Wismut cohort, to consider every aspect of this quantification would be very difficult and time-consuming for some sources of uncertainty (see Section *3.2* on the concept for the quantification). In this situation, it is important to find a trade-off between the relevance of a specific source of uncertainty in risk estimation and the effort that it takes to quantify and to account for this source of uncertainty. Concerning the relevance of the different sources of uncertainty in the Wismut cohort, we can deduce the following main points considering the relevance of the different sources of uncertainty:

- In accordance with Küchenhoff et al. (2018, p. 114), we have seen in this project that generalization error and parameter uncertainties are the most relevant sources of uncertainty and it is very important to account for these sources of uncertainty in risk estimation. They do not only occur for a large proportion of person work years in the cohort, but due to the fact that both sources of uncertainty contain classical measurement error components which are shared for several workers of the same object and also potentially for several subsequent years of exposure, they can have a very large impact on risk estimation (Hoffmann et al., 2018b).
- A parameter which deserves special attention is the proportion of the cumulative mined area in the previous years p. Indeed, there was a single estimation of this value p of 20%, which affects all years and miners belonging to M1a (11% PPY) in the same way and Lehmann et al. (1998) underlines the difficulty in the determination of this parameter.
- In accordance with Küchenhoff et al. (2018), assignment error appears to be of medium importance in the Wismut cohort. While it affects all exposure estimates in the cohort, it can be considered to only marginally affect risk estimation because of its pure unshared Berkson nature. To determine its exact influence on risk estimation, it is advisable to assess the impact of this type of error on risk estimation on simulated data with the exposure characteristics of the Wismut cohort.
- Contrary to Küchenhoff et al. (2018, p. 114), we think that transfer errors and experts' evaluation errors are of major importance in the Wismut cohort: Concerning transfer errors, we would suggest to consider these errors as special types of parameter uncertainties which can have a large influence on risk estimation. Concerning experts' evaluation error, we would characterize this source of uncertainty as classical measurement error as it can be argued that experts can be seen as very imprecise measurement devices. Moreover, we expect that far more exposure values are affected by expert estimation than assumed by Küchenhoff et al. (2018). In our view, experts' evaluation error is possibly present in a substantial proportion of person work years in the cohort. Based on the current information available on this type of error, which comes from the work of Allodji et al. (2012a), we would conclude in accordance with Küchenhoff et al. (2018) that there is a high degree of uncertainty in this expert estimation.

Table 3.1 gives for all uncertain parameters the values for the standard deviations with an explanation of how these values were obtained, as well as the respective measurement model where they occur and the assumed error type (additive or multiplicative and classical or Berkson).

We use the quantified value of the measurement error variance of the generalization error for object 009 Aue in 1961 and 1968 for all objects and years in M2 and M3, respectively. Since we are lacking any additional information to quantify the magnitude of error arising in the mean radon gas concentrations without mining activity at ground level and in a depth of 130m for objects 300 Lichtenberg in 1994/1995 ($\mathcal{C}_{Rn,0}(1994/95,300)$ and $\mathcal{C}_{Rn,130}(1994/95,300)$) occurring in M6, we also use the quantified variance value of the generalization error for object 009 Aue in 1961 for these two mean radon gas concentrations. For all transfer errors τ_e, τ_E and τ_{JEM}, the quantified values of the transfer error τ_{JEM} are used. All classical error components of parameter uncertainties that could not be quantified more accurately, were assigned the value 0.33 from the quantification of the evaluation factor in Section 3.4.2, whereas all Berkson error components of parameter uncertainties that could not be quantified more accurately, were assigned the value 0.69 from the quantification of the equilibrium factor.
The transfer of the error variances from one error component to another is defensible in this situation since we assume a multiplicative error structure where the magnitude of error remains comparable even if the uncertain parameter in itself shows a different magnitude.

Table 3.1: Uncertain parameters with the measurement models in which they occur, their assumed error types and the quantified standard deviations as well as the quantification approach. Due to lack of space, we abbreviated 'classical error standard deviation' to 'classical standard deviation' and 'Berkson error standard deviation' to 'Berkson standard deviation'.

parameter	measurement model	assumed measurement error type	standard deviation	quantification approach
$E(t,o,j)$	M1a, M1b, M2, M3, M4, M5a, M5b, M6	$U_{E,B}$: Berkson additive	13.95	quantified for 009 Aue 1961 assignment error
$\mathcal{E}(p_t,o)$	M4	$U_{\mathcal{E},c}$: classical multiplicative	0.936	Allodji et al. (2012b) for French cohort
$\mathcal{E}(o_0,j)$ $\mathcal{E}^*(p(t,o),o_0,j)$	M5a M5b	$U_{\mathcal{E},c}$: classical multiplicative	0.936	Allodji et al. (2012b) for French cohort
$\mathcal{C}_{Rn}(t,o)$	M2	$U_{\mathcal{C},c}$: classical additive	0.59	quantified for 009 Aue 1961 generalization error
$\mathcal{C}_{RDP}(t,o)$	M3	$U_{\mathcal{C},c}$: classical additive	0.03	quantified for 009 Aue 1968 generalization error
$\mathcal{C}_{Rn}(t_0(o_0(o)),o_0(o))$	M1a, M1b	$U_{\mathcal{C},c}$: classical additive	5.29	quantified for 009 Aue 1955 generalization error
$\mathcal{C}_{Rn}(1937/38,003)$	M1a	$U_{\mathcal{C},c}$: classical additive	6.56	quantified for 003 Schneeberg generalization error
$\mathcal{C}_{Rn,0}(1994/95,300)$, $\mathcal{C}_{Rn,130}(1994/95,300)$	M6	$U_{\mathcal{C},c}$: classical additive	0.59	quantified for 009 Aue 1961 generalization error
$\omega(p_t)$	M1a, M1b, M2, M3, M5a, M5b, M6	$U_{\omega,c}$: classical multiplicative $U_{\omega',B}$: Berkson multiplicative	0.04 0.69	Allodji et al. (2012a) for French cohort Berkson standard deviation as quantified for $\gamma(p_{to})$
$\varphi(o,j)$	M1a, M1b, M2, M3, M4, M5a, M5b, M6	$U_{\varphi,c}$: classical multiplicative $U_{\varphi',B}$: Berkson multiplicative	0.33 0.69	classical standard deviation as quantified for the evaluation factor Berkson standard deviation as quantified for $\gamma(p_{to})$
$\gamma(p_t,o)$	M1a, M1b, M2, M5b, M6	$U_{\gamma,c}$: classical multiplicative $U_{\gamma',B}$: Berkson multiplicative	0.23 0.69	quantified for classical error for $\gamma(p_{to})$ quantified for Berkson error for $\gamma(p_{to})$
$\varsigma(o)$	M3	$U_{\varsigma,c}$: classical multiplicative $U_{\varsigma',B}$: Berkson multiplicative	0.33 1.45	classical standard deviation as quantified for the evaluation factor quantified for Berkson error for $\varsigma(o)$
$\mathcal{b}(o)$	M1a	$U_{\mathcal{b},c}$: classical multiplicative $U_{\mathcal{b}',B}$: Berkson multiplicative	0.33 0.69	classical standard deviation as quantified for the evaluation factor Berkson standard deviation as quantified for $\gamma(p_{to})$
π	M1a	$U_{\pi,c}$: classical multiplicative $U_{\pi',B}$: Berkson multiplicative	0.79 0.69	quantified for classical error for π Berkson standard deviation as quantified for $\gamma(p_{to})$
$\epsilon(p_{to})$, $\epsilon_2(p_{to})$	M6	$U_{\epsilon,c}$: classical multiplicative $U_{\epsilon',B}$: Berkson multiplicative	0.33 0.69	quantified for classical error for the evaluation factor Berkson standard deviation as quantified for $\gamma(p_{to})$
$\zeta(o,j)$	M5a, M5b	$U_{\zeta,c}$: classical multiplicative $U_{\zeta',B}$: Berkson multiplicative	0.33 0.69	classical standard deviation as quantified for the evaluation factor Berkson standard deviation as quantified for $\gamma(p_{to})$
$\tau_e(o)$	M1a, M1b	$U_{\tau_e,c}$: classical multiplicative $U_{\tau_e',B}$: Berkson multiplicative	0.37 0.33	quantified for classical error for the transfer factor quantified for Berkson error for the transfer factor in M2
τ_E	M1a, M1b, M2, M3	$U_{\tau_E,c}$: classical multiplicative $U_{\tau_E',B}$: Berkson multiplicative	0.37 0.33 0.13	quantified for classical error for the transfer factor quantified for Berkson error for the transfer factor in M2 or M3, respectively
τ_{JEM}	M1a, M1b, M2, M3, M4, M5a, M5b, M6	$U_{\tau_{JEM},c}$: classical multiplicative $U_{\tau'_{JEM},B}$: Berkson multiplicative	0.37 0.33 0.13 0.18	quantified for classical error for the transfer factor quantified for Berkson error for the transfer factor in M2, M3 or M4/M5a/M5b, respectively

4 Developing a Bayesian hierarchical approach to correct for measurement error in the Wismut cohort

4.1 Potential levels of complexity to describe exposure uncertainty in cohorts of uranium miners

When accounting for exposure uncertainty in the Wismut cohort, it is advisable to consider different levels of complexity. As described in Chapter *2*, the exposure characteristics in the Wismut cohort are highly heterogeneous and depend on the type of object (e.g. processing companies, surface areas, etc.) and the exposure assessment period, yielding a complex overall measurement error structure. A first possibility to reduce the complexity of the method is to reduce the cohort to homogeneous groups of workers. For instance, the working condition hewer in object 009 Aue starting in 1960 represents a subsample of considerable size in the Wismut cohort with 50025 person work years corresponding to a proportion of person work years of 5.94%. By focusing on this subsample, we only need to account for the sources of exposure uncertainty present in measurement model M2 and M3 (see Sections *2.3* and 2.4). Moreover, when focusing on hewers in the analysis, we do not have to account for the measurement error in the activity weighting factor as hewer is the reference category. Overall, there are less unknown quantities intervening in exposure estimation for such a subsample and focussing on a subsample makes it easier to comprehend the error structure and the functioning of the approaches to correct for measurement error. In this vein, we can increase the complexity by subsequently enlarging the subgroup of miners. In a next step, we can for instance focus on hewers after 1960 in all underground mining objects and then on all workers after 1960 in underground mining objects. By then considering all workers after 1960 in all types of objects and then all workers in all exposure years, it is possible to gradually increase the complexity of the analysis by moving from subsamples with a manageable size to the full cohort. Similarly, we can focus in our analysis on the subsample of workers who were employed in processing companies (see measurement model M5a und M5b described in Section *2.6*) or on workers who were employed in surface areas affiliated to mining objects (see measurement model M4 described in Section *2.5*). While this procedure makes it somewhat easier to manage the complexity in the Wismut cohort, it entails a number of problems and difficulties. Firstly, when estimating the association between radon exposure and lung cancer mortality in the Wismut cohort, we have to consider the entire working history of a miner and there are only few miners who spent their entire working history in one working condition. For instance, it is likely that only a fraction of the miners who worked as hewers in object 009 after 1960 spent all their person work years in this working condition. More importantly, even when focusing on a homogeneous subsample, we are still faced with exposure data that are prone to complex structures of measurement error. Measurement model M2, for instance, which describes the uncertainty in the exposure assessment for a hewer after 1960 in object 009 still assumes a combination of generalization error and parameter uncertainties entailing both a Berkson and a classical component and which can be shared between workers, within workers or both. In this sense, focusing on a subsample cannot fundamentally reduce the complexity of the assumed measurement models.

An alternative approach to increase the complexity of the considered measurement error models is to consider simplifying assumptions which have been made in the literature when accounting for exposure uncertainty in occupational cohorts, and more specifically, in cohorts of uranium miners. Indeed, in the literature, one can find different approaches to describe exposure uncertainty in cohorts of uranium miners which vary in their level of complexity by making a number of simplifying assumptions. Table 4.1 provides an overview of these approaches with increasing complexity of the assumed measurement models.

On level 1, Küchenhoff et al. (2007) and Bender et al. (2005) who are interested in the effect of Berkson measurement error in Cox proportional hazards models and the generation of survival times following Cox proportional hazards models, respectively, use data of the Wismut cohort and make the simplifying assumptions that errors arise on the cumulative exposure values and that exposure

uncertainty can be described by homogeneous measurement error which is either of Berkson or of classical type.

Table 4.1: Different levels of complexity to describe exposure uncertainty in cohorts of uranium miners.

Level 1	- Unshared homogeneous measurement error of either classical or Berkson type - Measurement errors arise on the cumulative exposure	Küchenhoff et al. (2007) Bender et al. (2005)
Level 2	- Unshared heterogeneous measurement error - Group-level exposure assessment: Berkson - Individual exposure assessment: Classical - Measurement errors arise on yearly exposure values	Allodji et al. (2012a) Allodji et al. (2012b) Allodji et al. (2012c) Hoffmann et al. (2017)
Level 3	- Heterogeneous measurement error - Shared and unshared error components - Mixture of Berkson and classical error - Measurement errors arise on yearly exposure values	Hoffmann et al. (2018b)
Level 4	- Heterogeneous measurement error - Distinguish different measurement models according to object type - Distinguish different sources of uncertainty which are either of Berkson or classical type - Measurement errors arise on yearly exposure values	Küchenhoff et al. (2018)
Level 5	- Heterogeneous measurement error - Distinguish different measurement models according to object type - Distinguish different sources of uncertainty which can be mixtures of Berkson and classical type - Shared and unshared error components - Measurement errors arise on yearly exposure values	Chapter 2

On level 2, Allodji et al. (2012a,b,c) and Hoffmann et al. (2017) aim to correct the risk estimates in the French cohort of uranium miners and assume measurement error to arise on the yearly exposure values rather than on the cumulative exposure values as the radon exposure of an uranium miner is not a fixed point exposure at study entry, but ongoing during the miner's entire working career. By considering radon exposure to be time-dependent, they can model the measurement errors on their natural level of occurrence and distinguish three periods of exposure assessment. Allodji et al. (2012a) assess exposure uncertainty in these three periods of exposure assessment and identify various sources of exposure uncertainty ranging from the precision of the measurement device to the approximation of the equilibrium factor and the estimation of working time. The authors suggest to make the simplifying assumption that these different sources of uncertainty can be combined to derive a global measurement error which they assume either to be unshared Berkson error for group-level exposure assessment or unshared classical measurement error for periods of individual exposure assessment.

On level 3, Hoffmann et al. (2018b) argue against the simplifying assumption arising through the combination of different shared and unshared sources of uncertainty into a global unshared Berkson error as suggested by Allodji et al. (2012a) for the periods of group-level exposure assessment. In a simulation study based on the exposure characteristics of the French cohort of uranium miners, the authors show that a combination of shared and unshared Berkson and classical error components, which realistically reflects the sources of exposure uncertainty arising in a cohort of uranium miners, induces by far more bias in risk estimation and a more severe distortion of the exposure-response relationship than a pure unshared Berkson error component. When accounting for a combination of shared and unshared measurement error of classical and Berkson type in the French cohort of uranium miners, Hoffmann et al. (2018b) find that the excess hazard ratio (EHR) per 100 WLM is estimated to be 1.44 (95% credible interval: [0.66; 2.69]) compared to an estimated EHR per 100

WLM of 0.87 (95% credible interval: [0.49; 1.36]) and 0.90 (95% credible interval: [0.51; 1.41]) when ignoring measurement error and when solely accounting for unshared measurement error as suggested by Allodji et al. (2012a), respectively (results presented at the 40th Annual Conference of the International Society for Clinical Biostatistic in Leuven, Belgium in 2019).
On level 4, Küchenhoff et al. (2018) assess exposure uncertainty in the Wismut cohort and distinguish different object types with varying exposure assessment, which lead to different measurement models. Moreover, they distinguish the following sources of uncertainty that are either of Berkson or classical type: generalization error, parameter uncertainties, assignment error, transfer error, documentation error, experts' evaluation error, procedural measurement error and approximation error (Küchenhoff et al. 2018, p. 114).
Based on the extensive characterization of exposure uncertainty performed by Küchenhoff et al. (2018), on level 5 we can further increase the complexity by applying the rationale that Hoffmann et al. (2018b) applied to the French cohort of uranium miners to the Wismut cohort by assuming that the sources of uncertainty identified by Küchenhoff et al. (2018) can be mixtures of Berkson and classical type and that the dependencies arising in the exposure assessment of the Wismut cohort can be best described as shared measurement error (see Chapter *2*). While this combination leads to complex measurement error models that are challenging to implement in the Wismut cohort, it makes it possible to define measurement models that realistically reflect the exposure uncertainties arising in this cohort. Based on the results of Hoffmann et al. (2018b) it is likely that these complex measurement models induce more bias in risk estimation in the Wismut cohort than simple models that assume measurement errors to arise on the cumulative exposure during the entire working history of a miner (level 1) or models that assume unshared Berkson error for periods of group-exposure assessment (level 2). In order to assess the impact of these simplifying assumptions on the Wismut cohort, we will compare the impact of measurement error and the estimated risk when assuming measurement models of varying degree of complexity in the simulation study described in Chapter *5* and in our analyses on the real data of the Wismut cohort.

4.2 Requirements for a method to account for exposure uncertainty in the Wismut cohort

In order to account for exposure uncertainty in the association between radon progeny and lung cancer mortality in the Wismut cohort, it is necessary to choose a method which is applicable for proportional hazards models. It is advisable to choose a method that is flexible enough to be applied for different types of proportional hazards models, in particular including both Cox models of the more classical structure $h_i(t) = h_0(t)\exp(X_i\beta)$ and EHR models of the form $h_i(t) = h_0(t)(1 + X_i\beta)$, which are more commonly used in radiation epidemiology. Additionally, it is important to be able to account for different risk models when modelling the association between radon exposure and lung cancer mortality, for instance to account for effect modification by the exposure rate, which has been identified as an important effect modifying variable in cohorts of uranium miners in general (Lubin et al., 1995a; Vacquier et al., 2008) and in the Wismut cohort in particular (Walsh et al., 2010).
Based on the characterization of measurement error performed by Küchenhoff et al. (2018) and on the reflections exposed in the previous chapters, we see that the structure of measurement error in the Wismut cohort is very complex. As shown in the measurement models presented in Chapter *2*, there are several distinct sources of uncertainty, which may vary as a function of year of exposure, the object and the activity of a worker in the Wismut cohort. The measurement models we presented in Chapter *2* reveal that there are several uncertain quantities that simultaneously intervene in the estimation of radon exposure for a worker. These uncertain quantities are prone to distinct types of measurement error: For most activities, objects and years of exposure, we are confronted with one or several types of generalization error, which are assumed to be of classical type and shared between workers and with an assignment error which is of Berkson type and unshared. Additionally, there are several uncertain parameters intervening in exposure estimation for a given activity, exposure year and object and each uncertain parameter can be assumed to

include both a Berkson and classical measurement error component, which are shared between workers which were employed in the same object and year. In light of these facts, it is important to apply a method for measurement error correction in the Wismut cohort, which provides enough flexibility to realistically reflect the combination of the different types of measurement error and their heterogeneity arising in the exposure estimation in this cohort.

4.3 Approaches to account for measurement error in proportional hazards models

A common phenomenon when analyzing survival data through proportional hazards models is right-censoring, i.e. the failure time is not observed for every person i in the data set. Therefore, the outcome variable can be described by $Y_i = \min(T_i, C_i)$. Y_i is either the failure time T_i or the time until censoring C_i. An additional indicator δ_i, which is 1 if $T_i \leq C_i$ and 0 otherwise, provides the information of censoring or death for each person.
The most common approach to model the association between right-censored failure time and covariates (here the exposure to radon progeny) is the proportional hazards model. This model describes the instantaneous hazard rate $h_i(t)$ for person i by:

$$h_i(t) = h_0(t)\varphi(X_i, \beta)$$

where $h_0(t)$ is a baseline hazard which is equivalent to the instantaneous hazard rate for a person with no exposure to radon progeny. The function $\varphi(X_i, \beta)$ is used to model the hazard ratio as a function of the exposure X_i, and describes the hazard ratio between two persons where one is exposed with an exposure value of X_i and the other is not exposed. The original and widely used version of the proportional hazards model was proposed by Cox (1972), who assumes a log-linear relation, i.e. $\varphi(X_i, \beta) = \exp(X_i\beta)$. The exponential function ensures positivity for all real numbers. Therefore, no further constrains for β are necessary in contrast to other models like the Excess Relative Risk (ERR) model. A particularity of the proportional hazards model as proposed by Cox is that inference for this model can be conducted by relying on a partial, rather than a full likelihood (Cox, 1975). In this semi-parametric form of the proportional hazards model, no functional formulation for $h_0(t)$ is necessary.
A problem occurs if the exposure is prone to measurement error. Then it is not possible to formulate the hazard function as a distinct product of the baseline hazard which is only dependent on time t and a term that depends only on the data and parameters to be estimated (Prentice, 1982; Buzas, 1998; Yi and Lawless, 2007). Moreover, similar to logistic regression, measurement errors (classical as well as Berkson) can lead to a bias in the risk estimates in proportional hazards models (Kim et al., 2006; Küchenhoff et al., 2007). In the case of a single covariate with non-differential additive classical measurement error following a normal distribution and rare failure events, Hughes (1993) shows that the bias is approximately equal to the bias in a linear regression with a continuous outcome variable, i.e. the estimated β is attenuated by the factor $\sigma_x^2/(\sigma_x^2 + \sigma_u^2)$ where σ_u^2 is the the error variance. For a non-differential normally distributed additive classical error, Keogh et al. (2012) showed in a simulation study that non-linear associations between exposure and outcome appear more linear.
To conduct statistical inference taking measurement error into account one starts, like in a standard approach, specifying a response model $f(Y|X, \theta)$. This model, also often denoted disease model in epidemiology, formulates the relation between the outcome Y_i and the exposure covariate X_i (Richardson and Gilks, 1993a,b; Buzas et al., 2014). To account for measurement error, one additionally has to specify a measurement model. In the case of a classical error, this model describes the distribution of the observed exposure conditional on its true value. For a Berkson error, it is the other way around, i.e. the distribution of the true value given the observed. Depending on the approach to correct for measurement error, it is necessary to specify an additional parametric model which describes the probability distribution of true (and unobserved) exposure in the general population. For these called structural methods, one must therefore assign a known family of distributions to X, also known as exposure model (Congdon, 2006). In functional methods, on the

other hand, this is not necessary because the exposure is either seen as unknown and fixed or it can be seen as random, but with minimal assumptions concerning the distribution to be assumed (Carroll et al., 2006). In the literature, there is evidence that structural methods outperform functional methods concerning the efficiency if the exposure model is specified correctly. If this is not the case, they may produce erroneous results (Yi et al., 2015; Buzas et al., 2014; Schafer and Purdy, 1996; Küchenhoff and Carroll, 1997; Carroll et al., 2004; Messer and Natarajan, 2008). An overview of different methods to account for measurement error will be presented as follows. Firstly, the two most popular functional approaches, regression calibration and simulation extrapolation (SIMEX), will be introduced. Afterwards, two structural approaches will be described: the likelihood-based and the Bayesian approach.

4.3.1 Regression calibration

Regression calibration is a very popular technique and easy to understand. The basic idea is to substitute the unobserved true exposure by a predicted value from a calibration function. The calibration function predicts the expected value of X_i given the observed data point Z_i and other covariates V_i. Standard inference is then conducted with the replaced values which yields an unbiased estimate. The method has in general three steps (Carroll et al., 2006):

1. Estimation of the calibration function $E(X|Z,V)$;
2. Fit the disease model using $E(X|Z,V)$ instead of Z;
3. Calculate valid standard errors to account for the uncertainty in the first step.

To calculate the expected value of X, it is in general assumed that $E(X|Z,V)$ is a function of Z and other covariates V. An example would be to assume a linear connection. The result would be a linear regression model:

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

To use a model like this, a validation sample is needed where the true and observed covariates are measured (Thiébaut et al., 2007; Murad et al., 2016). This means that some true values of X must be at hand. If this is not the case, other options exist. For example, one could use an unbiased instrument T which is available for a subset of X and regress in step 1 T on Z and V. Another option would be to use replicate data. To use this approach, k_i replicates of each X_i must be available. Then it is possible to calculate the covariance matrix of the errors and use a linear approximation of the expectation of the true exposure given the error-prone Z and the replicates.
Regression calibration is a straightforward method. However, as stated by Carroll et al. (2006), *'No simple approximation can always be accurate'*, the method has its limitations. As stated above, some additional information like a calibration sample, repeated measurements or an instrumental variable is necessary. Moreover, regression calibration could lead to biased estimates. In linear models, regression calibration results in an unbiased estimate, though this does not hold for non-linear models like logistic or Cox regression (Augustin et al., 2008; Liao et al., 2011; Bartlett and Keogh, 2016; Althubaiti and Donev, 2016). Wang et al. (1997) provide the asymptotic theory for regression calibration in proportional hazards models. They also show that regression calibration, despite the inconsistency, yields a good finite sample performance for classical measurement error and non-time-varying covariates in the presence of a considerable validation sample. However, in complex settings, the estimators calculated by regression calibration can be even more biased than using a simple estimator which ignores the measurement error (Küchenhoff and Carroll, 1997).
In a last step in regression calibration, it is needed to adjust the standard errors of the estimates. The method uses two disjoint steps. In the first step, it calculates the conditional expectations of the true exposure given the error-prone measurement and other covariates. In the second step, these

expectations are used as a plug-in estimate for the true values X_i. This means that the uncertainty stemming from the estimation in the first step will be ignored. Therefore, one will get wrong standard errors and confidence intervals. Several authors proposed to apply bootstrap techniques, to remedy this problem (Carroll et al., 1996; Augustin et al., 2008; Spiegelman et al., 2011; Buonaccorsi et al., 2015). As an alternative, it is possible to use sandwich methods.
Allodji et al. (2012c) implement two variants of regression calibration, which he refers to as "substitution method" and "estimation calibration method", in a simulation study to study their performance when accounting for multiplicative exposure measurement error when analyzing the association between radon exposure and lung cancer mortality in the French cohort of uranium miners through Poisson regression.

4.3.2 Simulation extrapolation (SIMEX)

Simulation extrapolation (SIMEX) was originally proposed by Cook and Stefanski (1994). The basic idea is to model a functional relation between the estimated coefficient in a model and the measurement error. SIMEX is, as regression calibration, only approximately unbiased. Carroll et al. (1996) and Carroll and Stefanski (1997) provide the asymptotic theory behind the method (assuming classical additive normal distributed error). SIMEX has two different parts. The first part is a simulation step. It is required that the measurement error variance σ_u^2 is known or estimated well enough. Define for $\lambda_{sim} \geq 0$:

$$Z_{b,i}(\lambda_{sim}) = Z_i + \sqrt{\lambda_{sim}}\, U_{b,i}, \quad i = 1, \dots, n, \quad b = 1, \dots, B$$

where $U_{b,i}$ follows a normal distribution with zero mean. One can draw pseudo-random numbers since the variance σ_u^2 of $U_{b,i}$ is known. The term $(1 + \lambda_{sim})$ inflates the variance of the error term by a multiplicative factor. For $\lambda_{sim} = 0$, the generated errors get multiplied by a factor of zero. Hence, only the observed exposure remains. For $\lambda_{sim} \to -1$, the mean squared error (MSE) of $Z_{b,i}(\lambda_{sim})$, i.e. $\text{MSE}(Z_{b,i}(\lambda_{sim})) = E([Z_{b,i}(\lambda_{sim}) - X_i]^2 | X_i)$, converges to zero, which can be seen as a key property (Carroll et al., 2006). For the simulation part, one simulates B times $Z_{b,i}(\lambda_{sim})$ and conducts the naive estimation procedure for a fixed value of λ_{sim} which results in B estimates $\hat{\theta}_b(\lambda_{sim})$. Afterwards one calculates the mean over all estimates:

$$\hat{\theta}(\lambda_{sim}) = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b\,(\lambda_{sim}).$$

This is done for a fine grid of k values $\lambda_{sim} \geq 0$. Since one has to sample and estimate a model $B \times k$ times, the simulation step can be computationally expensive depending on the situation. After the simulation step, an extrapolation step follows. Here, the values of $\hat{\theta}(\lambda_{sim})$ get extrapolated back to $\lambda_{sim} = -1$. This leads to the simex estimator $\hat{\theta}_{\text{simex}}$. The performance of this estimator may heavily depend on the considered extrapolation function (Misumi et al., 2018). To get valid standard errors, it is again possible to use bootstrap techniques or a sandwich estimator (Cook and Stefanski, 1994; Carroll et al., 1996, 2006).
Several extensions for SIMEX for more or less special cases were developed. Nolte (2007) developed the M-SIMEX which can be used to account for multiplicative measurement error without using a log-transformation to get an additive error structure. Oh et al. (2018) used a SIMEX algorithm to correct for a multiplicative measurement error in the failure time outcome. In the field of radiation epidemiology, Masiuk et al. (2016) propose a version of SIMEX to account for classical additive and Berkson multiplicative errors in the association between thyroid doses caused by the Chernobyl accident and thyroid cancer in a cohort study in Ukraine. Allodji et al. (2012c) assess the performance of SIMEX when accounting for multiplicative Berkson and classical measurement error in Poisson regression in their simulation study based on the French cohort of uranium miners and Allodji et al. (2015) and Misumi et al. (2018) apply SIMEX when studying the association between radiation doses

and cancer incidence in the Life Span Study (LSS) of atomic bomb survivors in Hiroshima and Nagasaki through Poisson regression.

4.3.3 Likelihood-based approaches

After the presentation of two functional methods we switch to structural methods. These methods have the advantage that they can be applied to more general problems, lead to fully consistent estimators and more efficiency, i.e. smaller standard errors (Schafer and Purdy, 1996; Küchenhoff and Carroll, 1997; Messer and Natarajan, 2008). The resulting likelihood-based confidence intervals are more reliable than intervals obtained from SIMEX or regression calibration (Buzas et al., 2014), in particular for non-linear models (Carroll et al., 2006; Higdon and Schafer, 2001). These advantages are bought by stronger assumptions about the distribution of the exposure X (Carroll et al., 2006). Therefore, the method relies heavily on a correct specification of the exposure model (for classical measurement error) (Li and Lin, 2003; Gustafson, 2004). If this model is misspecified, the likelihood approach may be biased (see e.g. Li and Lin (2003)).
To account for measurement error in a likelihood-based approach, one has to combine the disease, the measurement and the exposure model (for classical error) to a joint likelihood. In the case of a classical measurement error, the likelihood can be formulated as

$$f(Y,Z|\theta) = \int f(Y,Z,X|\theta)dx = \int f(y|X,\theta)f(Z|X,\theta)f(X,\theta)dx$$

and for Berkson error with a modification:

$$f(Y,Z|\theta) = \int f(Y,Z,X|\theta)dx = \int f(y|X,\theta)f(X|Z,\theta)dx$$

To obtain $f(Y,Z|\theta)$, one has to evaluate an integral. For some cases, this can be done analytically, e.g. if all distributions are normal (Schafer and Purdy, 1996). In other cases, reasonable approximations exist. However, for most problems numerical methods must be used which leads to a high computational burden (Hu et al., 1998; Messer and Natarajan, 2008; Guolo and Brazzale, 2008; Buonaccorsi, 2010; Torabi, 2013).

4.3.4 Bayesian approaches

Starting from the likelihood-based approach, the Bayesian approach is not far away. As before, one starts by formulating a disease, a measurement and possibly an exposure model (for classical error). The difference lies in the view of parameters: In the likelihood approach, these parameters are unknown but fixed. In the Bayesian approach, they are assumed to be random. Therefore, one has to assign a *prior distribution* to each parameter. This could be an advantage in some situations: One can easily incorporate additional information about the parameters in a natural way. Moreover, the Bayesian approach gives an easy opportunity to take external information into account, e.g. from other studies (Carroll et al., 2006). If no additional information is available, one can use non-informative or flat priors, even though this may lead to non-identifiability issues in particular cases.

As described before, the integral in the likelihood-based approach could lead to a heavy computational burden. From a Bayesian point of view, the standard method of integrating over unknown quantities is the Markov chain Monte Carlo (MCMC) algorithm. Another possibility is Integrated Nested Laplace Approximation (INLA). Hereafter, both methods are briefly described.
The general aim of Bayesian inference is to connect the information from the data with the information from the priors to obtain the *posterior distribution*. An analytical calculation of the posterior distribution is only possible in very simple cases. But with the development of MCMC, in particular of the Metropolis-Hastings (MH) algorithm and the Gibbs sampler, it was possible to do inference with virtually any complex model (Metropolis et al., 1953; Hastings, 1970; Geman and Geman, 1984; Brooks et al., 2011). MCMC algorithms are numerical methods to sample from the posterior distribution. The algorithm creates a Markov chain which converges to the posterior distribution as stationary distribution. After convergence to the stationary distribution, the resulting values can be seen as draws from the posterior. If the chain runs long enough, the posterior is

sufficiently approximated. MCMC methods provide a rather general framework doing inference. This gives more freedom to focus on the development of accurate and realistic models without thinking too much about its computational feasibility in the first place. Richardson and Gilks (1993a) introduce a framework on how measurement models can be incorporated in a Bayesian manner by using Gibbs sampling to obtain posterior distributions. See for example Bartell et al. (2017) for a Bayesian approach to incorporate external information and Berkson error into an epidemiologic exposure-response analyses.
MCMC could approximate the posterior arbitrarily exact if computational resources and time would be infinite. However, in reality we want to obtain parameter estimates in finite time and MCMC sampling can become very time-consuming, especially in cases where the space of latent variables is very high-dimensional. With this in mind, Rue et al. (2009) propose INLA for structured additive regression models. This method approximates the marginal posteriors of the parameters of interest in a non parametric way. However, INLA relies heavily on normally distributed priors concerning the unknown quantities. Martins and Rue (2012) further developed the method for distributions which are not directly normal but near-normal (see Martins and Rue (2012) for more details). Muff et al. (2015) developed the INLA approach further to deal with measurement errors (classical and Berkson) and Muff et al. (2017) used INLA to deal with a combination of Berkson and classical measurement error with survival data. However, a quasi-normal assumption is always mandatory which limits the application to general problems.
The combination of the Bayesian hierarchical approach with MCMC provide greater flexibility than the structural approaches and likelihood based approaches. Hoffmann et al. (2017) use a Bayesian hierarchical approach to account for unshared heteroscedastic Berkson and classical measurement error when analyzing the association between radon exposure and lung cancer mortality in the French cohort of uranium miners through proportional hazards models. After studying the impact of more complex error structures with Berkson and classical measurement error components which can be shared and unshared in this cohort, Hoffmann et al. (2018b) suggest to use a Bayesian hierarchical approach to account for these complex error structures.

4.4 A Bayesian hierarchical approach to account for measurement error

As described in Sections *4.3.3* and *4.3.4*, Bayesian hierarchical approaches are very flexible when it comes to the correction of measurement error. With the complex error structures in mind as presented in Chapter *2*, hierarchical models allow to distinguish various levels of information (Greenland, 2000). Thus, it is possible to design a large and complex model by defining a disease model, a measurement model and (in case of classical measurement error) an exposure model and linking them by conditional independence assumptions (Richardson, 1996; Richardson et al., 2002; Ferrari et al., 2008). In contrast to many frequentist approaches, the definition of complex disease models including proportional hazards models to estimate EHRs is straightforward in a Bayesian hierarchical approach. Hereafter, we first describe the approach for very simple measurement error models (Berkson and classical homoscedastic) and then for the more complex and more realistic measurement model M2 (as defined in Section *2.3*).

4.4.1 The disease model

We model lung cancer mortality in the Wismut cohort by assuming a disease model, which describes the association between the right-censored variable time until death by lung cancer Y_i of miner $i, i \in \{1, \dots, n\}$ and the true cumulative radon exposure of miner i $X_i^{cum}(t)$ at time t. Traditionally, the association between radiation exposure and cancer mortality is described through grouped Poisson regression models in radiation epidemiology. It can be shown that in the case of a piecewise exponential survival distribution and categorical covariates, using a grouped Poisson regression modelling approach is equivalent to the analysis of survival times via proportional hazards models (Laird and Olivier, 1981). However, this approach involves the categorisation of continuous exposure

values and leads to the loss of information on the variance in each stratum as exposure values are typically averaged over the person-years in a stratum. More importantly, it is known in the literature that the categorisation of a continuous covariate with non-differential measurement error will in many cases lead to differential misclassification (Flegal et al., 1991). Finally, a Poisson regression modelling approach would impede the modelling of the individual. For these reasons, we chose to describe the association between radon exposure and lung cancer mortality in the Wismut cohort through proportional hazards models. In these models, the instantaneous hazard rate of death by lung cancer of miner i at age t, $h_i(t)$ is modeled through the following EHR model:

$$h_i(t) = h_0(t)(1 + \beta X_i^{\text{cum}}(t)).$$

To account for the complex patterns of exposure uncertainty in the Wismut cohort, we combine this disease model with different measurement models, which describe different types and magnitudes of error according to period of exposure and the type of objects (underground objects, open pit objects and surface areas associated to mining objects) in order to reflect changes in the methods of exposure assessment. In order to avoid the so-called time-dependent bias in survival analysis which arises whenever the future exposure status of an individual is analysed as being known from the beginning of follow-up (Walraven et al., 2004; Beyersmann et al., 2008; Wolkewitz et al., 2012; Barnett et al., 2011), we partition the follow-up time of the miners in the Wismut cohort in intervals at which the cumulative exposure was constant. By doing so, we create additional rows in the data set that we will refer to as "pseudo miners" where each time period at which cumulative exposure of a miner remained constant is represented as an independent row in the data set (Therneau and Grambsch, 2000; Therneau and Crowson, 2013). This data format is also referred to as long format as opposed to a wide format in which each worker is represented as one row and the information on yearly exposure values are given through additional columns.

4.4.2 Incorporating a single measurement error

In this section, we describe how a single homoscedastic measurement error (classical or Berkson) can be modeled via a Bayesian approach. To incorporate a classical measurement error, it is necessary to define three submodels: a disease model, a measurement model and an exposure model (Richardson and Gilks, 1993a). The models can be written in the following general forms:

$$\begin{aligned} &[Y|X, \theta_d] && \text{(disease model)} \\ &[Z|X, \theta_m] && \text{(measurement model)} \\ &[X|\theta_e] && \text{(exposure model)} \end{aligned}$$

where Y is the outcome, X is the true and unknown exposure and Z an error-prone version of X. θ_d, θ_m and θ_e are additional model parameters for each submodel.
In a Bayesian hierarchical framework, these models can be linked by using conditional independence assumptions. We are interested in the joint posterior distribution $[\boldsymbol{\theta}|Y, Z]$ with $\boldsymbol{\theta} = (\theta_d, \theta_m, \theta_e)$. The joint posterior distribution is the conditional distribution of the unknown parameters of interest $\boldsymbol{\theta}$ (and possibly the latent exposure X) given the observed data, i.e. the outcome Y and the error-prone version of the exposure Z. The corresponding joint posterior distribution is given by:

$$[\boldsymbol{\theta}, \mathrm{X}|Y, Z] \propto [\theta_d][\theta_m][\theta_e][X|\theta_e][Z|X, \theta_m][Y|X, \theta_d]$$

where the first three terms are the prior distributions for the unknown parameters in the different models.
When the aim is to describe a Berkson error structure, only two submodels have to be specified:

$$\begin{aligned} &[Y|X, \theta_d] && \text{(disease model)} \\ &[X|Z, \theta_m] && \text{(measurement model)} \end{aligned}$$

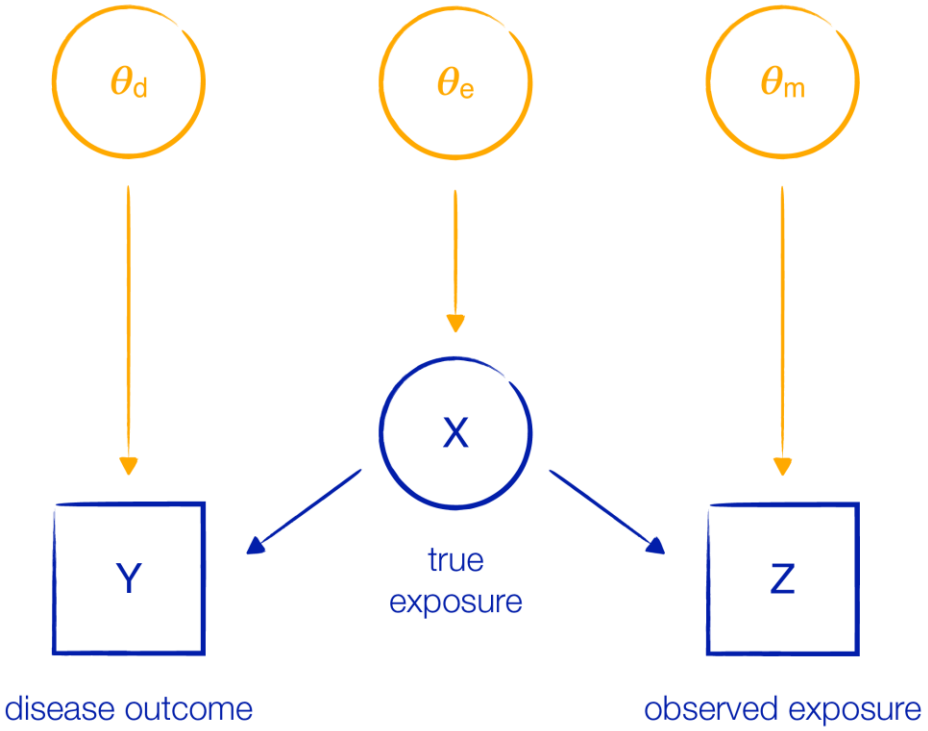

Figure 4.1: DAG for a single time-invariant classical measurement error.

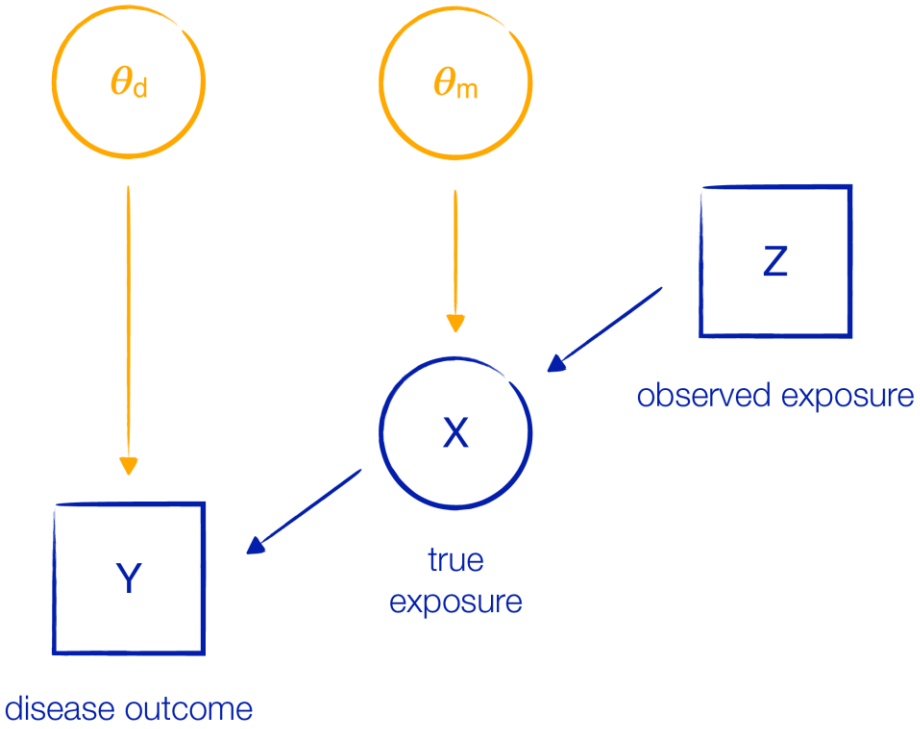

Figure 4.2: DAG for a single time-invariant measurement error of Berkson type.

It is not necessary to formulate an exposure model itself since the distribution of X is implicitly defined in the measurement model. For an illustration of the different models, it is possible, to represent them as a DAG (Jordan, 2004). Figure *4.1* shows the DAG for a classical measurement model and Figure *4.2* for Berkson error.
Since the true exposure X is not observable, the posterior distribution is derived by integrating over X. In this vein, the posterior is given by

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

Due to the integral over the unknown X, a closed form can be derived only in very simple cases and the integral could be approximated numerically. However, since the integral is often high-dimensional and numerical integration scales very poorly due to the curse of dimensionality (Hastie et al., 2009), it is necessary to use another approach. MCMC techniques provide enough flexibility and scale much better with the dimension of X and are therefore a good choice. Section *4.5* provides a more comprehensive description of the used technique. Since MCMC can be easily generalized to more complex models, we continue with more complex error structures.

4.4.3 Incorporating more complex measurement error structures: M2

The error structures in the Wismut cohort are much more complex than the simple pure Berkson or classical measurement error models which were described in the previous section. Therefore, it is required to use more sophisticated models. By using the Bayesian hierarchical approach this can be done in a straightforward manner.

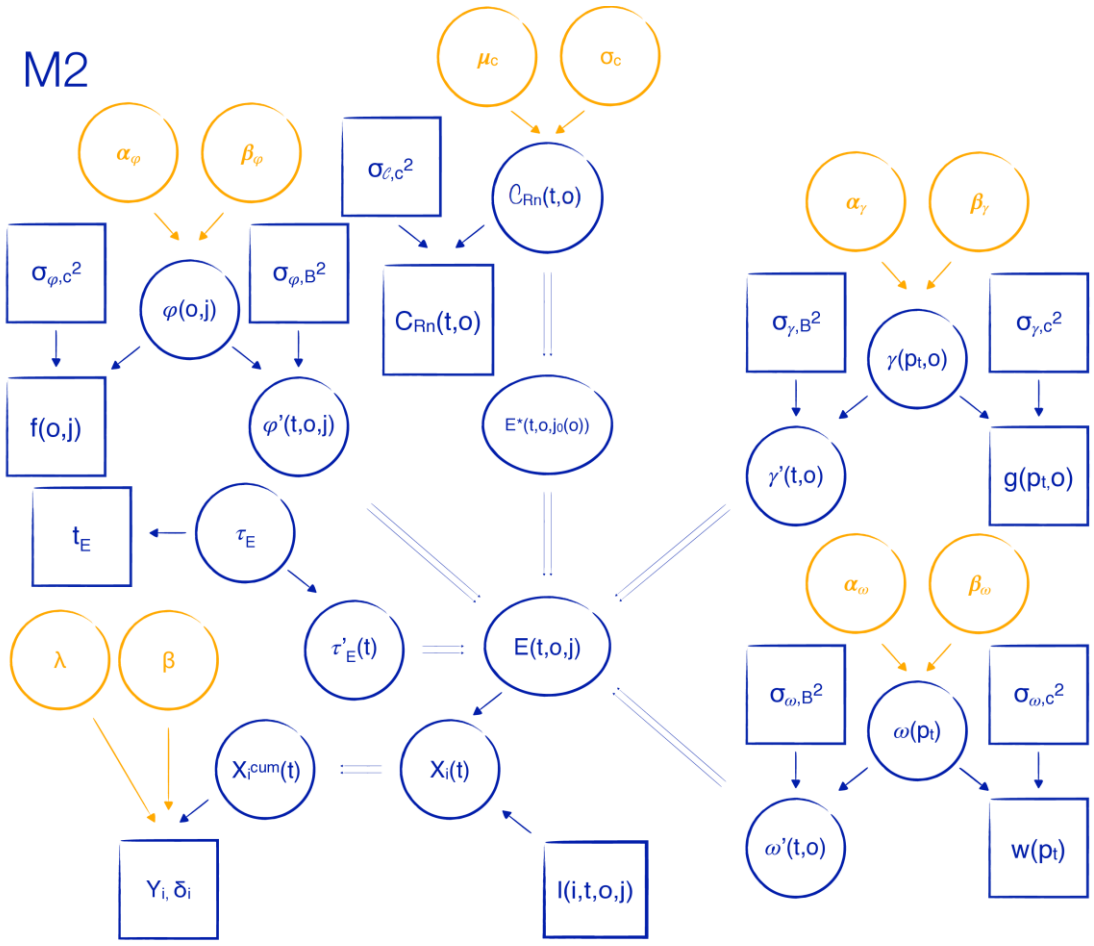

Figure 4.3: Hierarchical model combining a disease model with measurement model M2 to describe exposure uncertainty in underground mining objects in Saxony and Thuringia as well as in development objects in the second exposure assessment period. Due to the limited space and for a clearer presentation, no measurement error variances for t_E and $\tau_E(t)$ are shown here. Likewise, we omit the assignment error as we will omit it in the following presentation.

Since the Wismut cohort exhibits many different error structures depending on the methods of exposure assessment, it is not possible to use a single measurement model for all periods of exposure assessment. Hereafter, only measurement error model M2 will be described exemplarily. The full DAG is depicted in Figure *4.3*. Note that this is an extended version of the DAG from Section *2.3*: While the DAG in Figure 2.7 only depicted the disease and the measurement model, in Figure *4.3* we additionally show the exposure model describing the distribution of the various latent variables occurring in M2.

As can be seen in the DAG, the model consists of several parts where each error-prone quantity exhibits its own measurement and exposure model. The outcome Y_i for individual i depends on a time-varying covariate $X_i^{cum}(t)$. For more details on the challenges associated with the time-varying nature of $X_i^{cum}(t)$, see Section *4.6.2*. The model includes the following unknown quantities (which are all in circles in the DAG):

- Parameters to be estimated (marked in orange in the DAG):
  $\beta, \lambda, \alpha_\omega, \beta_\omega, \alpha_\gamma, \beta_\gamma, \alpha_\varphi, \beta_\varphi, \mu_\mathcal{C}, \sigma_\mathcal{C}$ where β and λ are the parameters of the disease model and the rest are parameters of the exposure model. We omit τ_E for brevity.
- Latent variables:
  $$\varphi(o,j), \varphi'(t,o,j), \gamma(p_t,o), \gamma'(t,o), \omega(p_t), \omega'(t,o), {\tau_E}'(t), \tau_E, \mathcal{C}_{Rn}(t,o), X_i(t)$$

We have two parameters for the exposure model of each of the factors (e.g. α_ω and β_ω for ω). These parameters depend on the assumed exposure distribution. In the case of a Beta distribution, these parameters are $\alpha_{(\cdot)}$ and $\beta_{(\cdot)}$ and $\mu_{(\cdot)}$ and $\sigma_{(\cdot)}$ for the normal and the lognormal distribution.
The joint posterior can be expressed as

$$\begin{aligned}
[\theta, X|\cdot] = \quad & [\beta][\lambda][\alpha_\omega][\beta_\omega][\alpha_\gamma][\beta_\gamma][\alpha_\varphi][\beta_\varphi][\mu_\mathcal{C}][\sigma_\mathcal{C}] \times \\
& \prod_{i,t} [\, Y_i|\lambda, \beta, X_i^{cum}(t)] \times \\
& \prod_{i,t} [\, X_i(t)|\mathcal{C}_{R_n}(t,o), \varphi'(t,o,j), \gamma'(t,o), \omega'(t,o), {\tau_E}'(t), l(i,t,o,j)] \times \\
& \prod_{t,o} [\, \omega'(t,o)|\sigma^2_{\omega',B}, \omega(p_t)] \prod_{p_t} [\, w(p_t)|\sigma^2_{\omega,c}, \omega(p_t)] \prod_{p_t} [\, \omega(p_t)|\alpha_\omega, \beta_\omega] \times \\
& \prod_{t,o} [\, \gamma'(t,o)|\sigma^2_{\gamma',B}, \gamma(p_t,o)] \prod_{p_t,o} [\, g(p_t,o)|\sigma^2_{\gamma,c}, \gamma(p_t,o)] \prod_{p_t,o} [\, \gamma(p_t,o)|\alpha_\gamma, \beta_\gamma] \times \\
& \prod_{t,o,j} [\, \varphi'(t,o,j)|\sigma^2_{\varphi',B}, \varphi(o,j)] \prod_{o,j} [\, f(o,j)|\sigma^2_{\varphi,c}, \varphi(o,j)] \prod_{o,j} [\, \varphi(o,j)|\alpha_\varphi, \beta_\varphi] \times \\
& \prod_{t,o} [\, \mathrm{C}_{R_n}(t,o)|\sigma^2_{\mathcal{C},c}, \mathcal{C}_{R_n}(t,o)] \prod_{t,o} [\, \mathcal{C}_{R_n}(t,o)|\mu_\mathcal{C}, \sigma_\mathcal{C}]
\end{aligned}$$

where we use the notation $[\theta, X|\cdot]$ to write the distribution of θ and the latent true exposure X conditioned on everything else. For brevity, we omit $l(i,t,o,j)$ and ${\tau_E}'(t)$ in the following when we refer to measurement model M2. We are therefore left with four constituent factors representing the latent true exposure. The joint posterior has no closed solution. Therefore, MCMC methods seem to be a proper way to conduct inference. The following two sections describe how MCMC works in general and how it is implemented in the context of a given error structure.

4.5 Conducting Bayesian inference through Markov Chain Monte Carlo (MCMC) methods

As described before, the posterior distribution cannot be derived in a closed form in the most practical settings. Therefore, it is necessary, to use computational methods for an adequate approximation. The possibly most flexible way to do this are MCMC methods. By using an MCMC algorithm, it is possible to draw samples from a complex distribution without knowing the

normalizing constant which requires in general the calculation of a potential high-dimensional integral (Brooks, 2003; Green et al., 2015).
MCMC algorithms generate a discrete-time Markov chain. A Markov chain is a collection of random variables S_t (states) where an index $t = 1,2, \ldots$ t,t+1, … defines a natural ordering, i.e. represents equally distant time points (note that we use a non-cursive to indicate the index of a Markov chain in contrast to t which indicatees the actual time domain). The variables satisfy the Markov property which states that the conditional distribution of the future state S_{t+1} only depends on the previous state S_{t+1}, but not on the states before S_t i.e. $[S_{t+1}|S_t, S_{t-1}, S_{t-2}, \ldots, S_1, S_0] = [S_{t+1}|S_t]$, where S_k refers to the state of the chain at time k and S_0 refers to the initial state of the chain. Under certain conditions, a Markov chain converges to a stationary distribution. See for instance Roberts and Smith (1994) or Robert and Casella (2004). The realizations from the chain can be seen as draws from this stationary distribution. In the context of Bayesian inference, this distribution is the joint posterior distribution of the model. MCMC methods provide a generic approach to conduct Bayesian inference and various algorithms exist to generate a Markov chain which exhibits the posterior as stationary distribution. The first algorithms stem from the statistical physics literature (Besag and Green, 1993; Metropolis, 1987; Landau and Binder, 2014; Robert and Casella, 2011), and research projects involving the development of the atomic bomb and the first nuclear reactors (Andrieu et al., 2004; Shonkwiler and Mendivil, 2009).
In contrast to many other numerical techniques, such as rejection sampling, in MCMC, after convergence every realization can be seen as a sample from the posterior distribution whereas in rejection sampling often multiple trials are necessary to obtain only one realization from the target distribution.
However, one has to face other problems. The generated samples from the Markov chain can exhibit a high autocorrelation. This means the iterative draws are highly correlated and the effective sample size gets reduced. Another point is the convergence behavior of the Markov chain to the target distribution. There are no diagnostic tools which guarantee that the chain has converged to its stationary distribution, but only detection tools showing divergence. The most common way was proposed by Gelman and Rubin (1992). Here, multiple Markov chains are generated with dispersed starting values. Afterwards, they are compared with respect to their variability within and between the chains. After dropping a sufficient high burn-in (ignoring the first values of the chain until convergence is achieved), a calculated ratio should be close to one. This also means that all chains visit the whole of the sample space of interest.
The by far most popular MCMC method is the MH algorithm, which will be presented in the following.

4.5.1 The Metropolis-Hastings algorithm

When conducting Bayesian inference via the MH algorithm, one has to specify a proposal distribution $[\boldsymbol{\theta}|\boldsymbol{\theta}_{t-1}]$ according to which new values $\boldsymbol{\theta}_{cand}$ are drawn given the current state $\boldsymbol{\theta}_{t-1}$. The full algorithm is given as follows.

Specify proposal distribution $[\boldsymbol{\theta}|\boldsymbol{\theta}_{t-1}]$
Initialize all variables $\boldsymbol{\theta_0}$

for *t = 1*: #iterations **do**

Generate a candidate $[\boldsymbol{\theta}_{cand}|\boldsymbol{\theta}_{t-1}]$

Calculate acceptance rate ρ:

$$\rho(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{cand}) = \min\left\{\frac{[\boldsymbol{\theta}_{cand}|X_{t-1}, Y]}{[\boldsymbol{\theta}_{t-1}|X_{t-1}, Y]}\frac{[\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_{cand}]}{[\boldsymbol{\theta}_{cand}|\boldsymbol{\theta}_{t-1}]}, 1\right\}$$

Accept $\boldsymbol{\theta}_{cand}$ as new state $\boldsymbol{\theta}_t$ with probability ρ

Otherwise set $\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1}$

end
Discard the first M iterations as burn-in.

As before, we use Z to denote the observed exposure and Y to denote the disease outcome. The algorithm proposes new values in each iteration t. These new values get accepted as a new state of the Markov chain with probability ρ. If the proposed value gets rejected, the new state is set to the old value of $\boldsymbol{\theta}$. For the calculation of ρ, it is necessary to calculate the product of two ratios: $[\boldsymbol{\theta}_{\text{cand}}|Y,Z]/[\boldsymbol{\theta}_{\text{t-1}}|Y,Z]$ and $[\boldsymbol{\theta}_{t-1}|\boldsymbol{\theta}_{\text{cand}}]/[\boldsymbol{\theta}_{\text{cand}}|\boldsymbol{\theta}_{\text{t-1}}]$. This product ensures, that the chain converges to the desired distribution. The acceptance rate (the proportion of accepted states of all proposed states) and therefore the successful design of the algorithm depends heavily on the appropriate choice of the proposal distribution. This proposal distribution must be known analytically (up to a unknown constant which cancels out in the ratio). Furthermore, it should be possible to propose values in a time-efficient way. In the case of a symmetric proposal distribution the latter part of the product collapses to one. Therefore, only the ratio of the posterior distribution evaluated at $\boldsymbol{\theta}_{\text{cand}}$ and $\boldsymbol{\theta}_{\text{t-1}}$ remains and the algorithm degenerates to the Metropolis algorithm.

The performance of the algorithm depends heavily on the acceptance rate. Therefore, it is reasonable, to monitor this rate. Moreover, it 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.
It is not necessary to update all components of $\boldsymbol{\theta}$ at once. Instead it can be convenient to split $\boldsymbol{\theta}$ in smaller components $(\theta_1, \theta_2, \dots)$ and to update these components one by one (Gilks et al., 1996; Andrieu et al., 2010).

4.5.2 Deriving estimators from the posterior distribution

After the algorithm generated a sufficient long Markov chain and after dropping an adequate number of initial samples as burn-in, the realizations can be seen as a large sample from the posterior distribution and represents therefore a good approximation. Deriving estimators for various quantities from this sample can be done in a straightforward manner. For example, for a point estimate one can choose the mean or the median as a robust alternative. Estimates for other quantities like the variance or a credible interval can also be substituted with their empirical analogues. For instance, the empirical 0.025 and 0.975 quantiles give a good estimate of the 95% credible interval of the posterior distribution.

4.6 Implementation of an efficient MCMC algorithm

After describing the general procedure of the MH algorithm, this concept will be used now to picture how concrete updates in the context of the Wismut cohort look like. In a first step, an exemplary update of the MH algorithm for a simple model as outlined in Section 4.4.1 will be described. In this simple case, we assume that neither the relation in the disease model nor the exposure variable itself is time-dependent. After this simplified example, we will describe the update for measurement model M2 from Section 4.4.2.

4.6.1 Update in a Bayesian hierarchical model with a single classical measurement error

Looking back to Section 4.4 shows that the parameters are given by $\boldsymbol{\theta} = (\theta_d, \theta_m, \theta_e)$. For ease of readability, the joint posterior is stated again:

$$[\boldsymbol{\theta}, \mathrm{X}|Y,Z] = [\theta_d][\theta_m][\theta_e][X|\theta_e][Z|X,\theta_m][Y|X,\theta_d]$$

As stated before, the update of $\boldsymbol{\theta} = (\theta_d, \theta_m, \theta_e)$ can be done one after another. To show how the update works in the context of the MH algorithm we assume to be at t-1 and we want do an update at time point t. Therefore, the state of the Markov chain before the update is given by

$$\boldsymbol{\theta}_{\text{t-1}} = (\theta_{d,\text{t-1}}, \theta_{m,\text{t-1}}, \theta_{e,\text{t-1}}).$$

Since X is not observable, it is necessary to treat it as a latent variable which is like the parameters an unknown quantity. Therefore, it must be updated in each iteration step of the algorithm as well as the parameters $\boldsymbol{\theta}$.

Update of the parameters of the disease model

For an update of $\theta_{d,\text{t-1}}$ to $\theta_{d,\text{t}}$, only the parts of the joint posterior are relevant, where $\theta_{d,t-1}$ is actually present. The rest of the parameters (θ_m, θ_e) do not influence θ_d meaning they can be seen as a constant and can therefore be ignored. Be doing so we are using the so-called full conditional of θ_d. This full conditional is given by

$$[\theta_d|X, Y] \propto [\theta_d][Y|X, \theta_d].$$

Therefore, after proposing a new value $\theta_{d,\text{cand}}$ given the current state $\theta_{d,t-1}$ would be accepted with probability

$$\rho(\theta_{d,\text{t-1}}, \theta_{\text{cand}}) = \min\left\{\frac{[\theta_{d,\text{cand}}|X_{\text{t-1}}, Y]}{[\theta_{d,\text{t-1}}|X_{\text{t-1}}, Y]}\frac{[\theta_{d,\text{t-1}}|\theta_{d,\text{cand}}]}{[\theta_{d,\text{cand}}|\theta_{d,\text{t-1}}]}, 1\right\}$$

where $X_{\text{t-1}}$ is the current state of the latent variable. The term $\frac{[\theta_{d,t-1}|\theta_{d,\text{cand}}]}{[\theta_{d,\text{cand}}|\theta_{d,\text{t-1}}]}$ is a ratio of two probabilities: the nominator is the likelihood of obtaining $\theta_{d,\text{t-1}}$ given $\theta_{d,\text{cand}}$. The denominator evaluates the same in the opposite direction. As stated before, this term is not necessary in a symmetric proposal distribution, because it cancels out. This circumstance can also be exploited in the implementation, to accelerate the calculation of the ratio. After calculation of ρ, the new state is set to $\theta_{d,t} = \theta_{d,\text{cand}}$ with probability ρ or $\theta_{d,t} = \theta_{d,t-1}$ with probability $1 - \rho$.
After updating the state from $\theta_{d,\text{t-1}}$ to $\theta_{d,\text{t}}$, this new state is used for the update of the other parameters (and the latent true exposure variable). Therefore, the new intermediate parameter vector is given by

$$\widetilde{\boldsymbol{\theta}} = (\theta_{d,\text{t}}, \theta_{m,\text{t-1}}, \theta_{e,\text{t-1}}).$$

This new state can again be used to update the other parts of $\widetilde{\boldsymbol{\theta}}$ until all updates are done. After updating the parameters to state $\boldsymbol{\theta}_{\text{t}}$, the latent true exposure variable X must be updated too. After a full update at time point t, this procedure can be repeated for t+1.

Update latent true exposure variable $\boldsymbol{X}$

The update for X can be done in the same manner. Mathematically, it can be treated as a parameter. The full conditional is given by:

$$[\text{X}|\boldsymbol{\theta}, Y, Z] \propto [X|\theta_e][Z, |X, \theta_m][Y|X, \theta_d]$$

For the update of X one needs again a proposal distribution to draw a candidate X_{cand} given $X_{\text{t-1}}$. The acceptance rate is calculated as follows:

$$\rho(X_{\text{t-1}}, X_{\text{cand}}) = \min\left\{\frac{[X_{\text{cand}}|\boldsymbol{\theta}_{\text{t}}, Y, Z]}{[X_{\text{t-1}}|\boldsymbol{\theta}_{\text{t}}, Y, Z]}\frac{[X_{\text{t-1}}|X_{\text{cand}}]}{[X_{\text{cand}}|X_{\text{t-1}}]}, 1\right\}$$

4.6.2 Update in a Bayesian hierarchical model with the complex measurement model M2 for $\boldsymbol{\beta}$

The updates for the more complex model M2 basically work the same way with the notable difference that the latent unknown exposure is a combination of different unknown factors (in measurement model M2, X is a product of these factors multiplied by 12). Therefore, instead of one latent variable X, each of the constituent parts ($\varphi'(t, o, j)$, $\gamma'(p_t, o)$, $\omega'(t, o)$ and $\mathcal{C}_{Rn}(t, o)$ for M2)

has to be updated in each iteration besides the other parameters. For simplicity, we slightly modified the uncertainty structure described for M2 in *2.3.2* by not including the uncertain quantity $\tau_E{}'$. We show how the update for one of the factors of the latent variable is made in Appendix *A 1* since we are faced with various technical challenges which we will elaborate in more detail there. Here, we show how the update for the parameter of major interest, β, works for the given individual exposure of every worker and year. For this update, we assume to have the latent exposure for each worker i and time t. The full conditional for β is given by:

$$[\beta|\lambda, X^{cum}] \propto [\beta] \prod_{i,t} [\, Y_i|\lambda, \beta, X_i^{cum}(t)]$$

The rest of the joint posterior is independent of β and can therefore be neglected. Given the current state β_{t-1}, a new state β_{cand} is drawn from the proposal distribution. This candidate will be accepted with probability

$$\rho(\beta_{\text{t-1}}, \beta_{\text{cand}}) = \min\left\{\frac{[\beta_{\text{cand}}|\lambda, X_{\text{t}}^{cum}]}{[\beta_{\text{t-1}}|\lambda, X_{\text{t}}^{cum}]} \frac{[\beta_{\text{t-1}}|\beta_{\text{cand}}]}{[\beta_{\text{cand}}|\beta_{\text{t-1}}]}, 1\right\}$$

Here we face some challanges: The disease model links the outcome Y to the cumulative exposure X^{cum} while in the measurement model and exposure model are defined in terms of uncumulative exposures. Hence, before the update of β and λ can be done, it is necessary to cumulate $X(t)$ for each person i. A small example shows how X^{cum} could look like:

$$X^{cum} = \begin{pmatrix} X_1(1) \\ X_1(1) + X_1(2) \\ X_1(1) + X_1(2) + X_1(3) \\ X_1(1) + X_1(2) + X_1(3) + X_1(4) \\ X_1(1) + X_1(2) + X_1(3) + X_1(4) + X_1(5) \\ X_2(1) \\ X_2(1) + X_2(2) \\ X_3(1) \\ X_3(1) + X_3(2) \\ X_3(1) + X_3(2) + X_3(3) \end{pmatrix}$$

Here the first person worked for five years, the second two and the third worked for three years. Therefore, in each iteration, after the update of the latent variables, it is necessary to cumulate $X(t)$ again to obtain the cumulative exposure vector X^{cum} .
The exposure is high-dimensional. Besides, there are many other unknown quantities, which are also high-dimensional. Despite the computational advantages of MCMC, this can lead to a high autocorrelation in the generated Markov chain. Therefore, a large number of iterations is required implying a high computational effort.

4.6.3 Implementation of the algorithm

To conduct Bayesian inference in the aforementioned manner, we choose to implement the MCMC algorithm in an object-oriented fashion to benefit from the greater flexibility through a modular design. This allows us to adjust parts of a module without having to modify the other parts of the code. Moreover, it allows to tweak the algorithm at specific positions, e.g. implementing an alternative solution for better numerical stability. Finally, an object-oriented implementation has the benefit that the program code is naturally divided into several functional files that are self-contained and show a clear dependence structure that can be visualized in a class diagramm. We also choose a test-driven development approach (Janzen and Saiedian, 2005). In this test-driven development approach, all functionalities of the algorithm are tested by unit tests to ensure their correctness. This leads to much less 'black-box errors' which are hard to detect. The object-oriented programming language *Python3* provides many modules, to meet our requirements:

- *Numpy* provides functionalities for fast computation, especially array-like data structures.
- `scipy.stats` provides various statistical methods, e.g. random number generators for various distributions.
- *pandas* was used for efficient data management.
- `scipy.sparse` implements objects and methods for sparse matrix objects which can speed up matrix operations by exploiting the fact that the majority of the matrix contains zeros and can therefore be ignored.
- *pandas* is used for efficient data management.
- `unittest` provides a good unit test framework.

An editable version of the code is provided.

4.6.4 Required input

Data frame

In order to run the algorithm for M2, the user has to provide a data frame with the information of the Wismut cohort in long format. This means that each miner is represented by one row for every year with the obtained radon exposure. Each row also has to contain the info whether the follow-up period of the miner ended in the considered year (either due to censoring or due to death of lung cancer). Table *4.2* shows all variables that need to be provided by the user and their names in the data frame, as well as their analogues in M2 (see the DAG in Figure *2.7*). In order to apply the algorithm to the Wismut cohort, information about the mean concentration of the radon gas measurements, the working time factor, the equilibrium factor and the activity weighting factor are needed. The necessary information for the working time and the equilibrium factors is available in Tables 9 and 10 of Küchenhoff et al. (2018, pp. 44-45) and the values for the activity weighting factor can be found in Table 2.3.1.1 in Lehmann et al. (1998, pp. 256–266). The working time and the equilibrium factors each consist of only a few different values depending on the time period and for each of the two factors a shared measurement error is assumed for all years with the same value. Therefore, the data frame needs variables indicating the respective periods for which the working time factor values and the equilibrium factor values are the same. In this sense, the variables 'w_period' and 'g_period' can be derived based on the information given in Tables 9 and 10 in Küchenhoff et al. (2018, pp. 44-45), respectively. The mean radon gas or radon progeny concentration measurements can be obtained from Lehmann et al. (1998) and Lehmann (2004). Note that the implementation of the algorithm does not require the user to specify the observed values of the exposure $Z_i(t)$ and the cumulative exposure $Z_i^{cum}(t)$ directly when correcting for measurement errors. Instead, the exposure of miner i at time t, $X_i(t)$, is derived from the constituent parts of $X_i(t)$ according to the formula of model M2, i.e. $\varphi'(t, o, j)$, $\gamma'(p_t, o)$, $\omega'(t, o)$ and $\mathcal{C}_{Rn}(t, o)$.

Due to the flexibility of the implemented algorithm, it is straightforward to apply the algorithm for other causes of death and for other exposures in the Wismut cohort. To do so, it is not necessary to change the algorithm but only to modify the input variables. If the measurement error characteristics are not the same for this exposure, the user can also modify the information provided through the uncertainty characteristics file (see next paragraph "Uncertainty characteristics"). In order to apply the algorithm to estimate the association between long-lived radionuclides and leukemia mortality, for instance, the user simply has to 1) define a measurement model describing how long-lived radionuclides were assessed in the cohort, 2) provide the names of the observed values of the uncertain quantities intervening in this measurement model and 3) define the variables stop and delta as a function of the time points during the history of the worker at which exposure to long-lived radionuclides changed and at the time at which the worker was either censored (delta = 0) or died of

leukemia (delta = 1). This example assumes, that the error structure for long-lived radionuclides is the same as in M2.

Table 4.2: Variables that need to be provided by the user.

variable	**M2**	**name in the data frame**
year	t	year
object	o	object
activity	j	activity
miner's age at the beginning of the time interval		start
miner's age at the end of the time interval		stop
indicator for death due to lung cancer	δ_i	delta
observed mean concentration of the radon gas measurements	$C_{Rn}(t, o)$	C_Rn_obs
observed working time factor	$w(p_t)$	w_classical
working time factor period		w_period
observed equilibrium factor	$g(p_t, o)$	g_classical
equilibrium factor period		g_period
observed activity weighting factor	$f(o, j)$	f_classical

Uncertainty characteristics

The error structure in each measurement model can be specified through a certain number of attributes, which we denote as uncertainty characteristics. The implemented algorithm requires a file which provides information on these uncertainty characteristics. The file has to be in the form of a python-dictionary. Therefore, it has to contain key-value-pairs representing the different characteristics of the error structure. In particular, the file has to take the form of a nested dictionary consisting of various sub-dictionaries. The first level takes as keys the model and its model specific names of uncertain factors as sub-keys for the sub-dictionary. With the term '(uncertain) factor' we refer to the constituent parts of a measurement model which are used to calculate the exposure X (for measurement model M2, these factors are φ, γ, ω and $\mathcal{C}_{Rn}$). Factors which are used over different models are not specified for each model separately. They have to be specified in the first level with its actual factor name and in its second level with its symbolic name as represented in the DAG. For instance, the working time factor is used in different measurement models besides M2 and is therefore separated from M2 in the first level. It has its own key 'working_time' on the first level and 'omega' as key on the second level. The mean concentration of the radon gas measurements C_{Rn} is only used in M2 and therefore in the sub-dictionary 'M2'. The last level specifies the actual measurement error structure of the uncertain factor as a third-level-dictionary. Here, each key indicates the exact specification for the factor as follows:

- classical_error: A dictionary containing the keys 'sd' (standard deviation of the error), 'structure' (either multiplicative or additive), 'proposal_sd' (the initial standard deviation to make proposals)

- Berkson_error: A dictionary containing the keys 'sd' (standard deviation of the error), 'structure' (either multiplicative or additive), 'proposal_sd' (the initial standard deviation to make proposals)
- exposure_model_distribution: A string defining the distribution of the exposure model
- exposure_model_parameters: A dictionary defining the parameter values for the exposure distribution
- exposure_model_truncation: A dictionary with the keys 'lower' and 'upper' defining the range of possible values of the uncertain factor
- mapping_identifier_classical: A list of strings defining the domain the error depends on
- mapping_identifier_Berkson: A list of strings defining the domain the error depends on. These strings are the variables names of the data frame. For example, the information *['year', 'object']* would suggest individual errors for each year and object in the data over all relevant rows.
- name_obs_values: The name of the variable from the input data frame which contains the observed values of the uncertain parameter.

The user has to specify the uncertainty characteristics for all relevant uncertain factors according to the used measurement model formula. In the case of M2 these are φ, γ, ω and $\mathcal{C}_{Rn}$. Since the first three of them are also part of other measurement models (for example M3), they are not part of the key 'M2' and only $\mathcal{C}_{Rn}$ remains in 'M2'.
To get a more realistic idea on how the uncertainty characteristics file should look like, a partial file for the uncertain factors $\mathcal{C}_{R_n}$ and ω for measurement model M2 is presented below. This file is also used in the simulation study for Scenario S2 and S3 (see Chapter 5).

```
uncertainty_characteristics = {
      'M2':{
            'C_Rn': {
                  'classical_error':{'sd': 0.59,
                                     'structure': 'additive',
                                     'proposal_sd': 0.1},
                  'Berkson_error':{'sd': 0},
                  'exposure_model_distribution': 'norm',
                  'exposure_model_parameters': {'mu': 6,
                                                'sigma': 8},
                  'exposure_model_truncation': {'lower': 1e-10},
                  'mapping_identifier_classical' : ['year',
                                                    'object']
                  'name_obs_values': 'C_Rn_obs'
            },
      },
      'working_time':{
            'omega': {
                  'classical_error':{'sd': 0.33,
                                     'structure': 'multiplicative',
                                     'proposal_sd': 0.1},
                  'Berkson_error':{'sd': 0.69,
                                   'structure': 'multiplicative',
                                   'proposal_sd': 0.1},
                  'exposure_model_distribution': 'beta',
                  'exposure_model_parameters': {'alpha': 3,
                                                'beta': 3},
                  'exposure_model_truncation': {'lower': 0.88,
                                                'upper': 1.2},
                  'mapping_identifier_classical': ['w_period']
```

```
                'mapping_identifier_Berkson': ['year',
                                               'object']
                'name_obs_values': 'w_classical'
            },
        },
        'activity':{
            'phi': { …},
        },
        'equilibrium':{
            'gamma': { …},
        },
}
```

The factor $\mathcal{C}_{Rn}$ is only part of M2 while ω is used in other models as well. Therefore $\mathcal{C}_{Rn}$ is encapsulated in the sub-dictionary 'M2' and ω in its own sub-dictionary 'working_time'. In the simulation study in Chapter *5*, $\mathcal{C}_{Rn}$ only has an additive normal classical error with a standard deviation of 0.59. Hence, the specification for $\mathcal{C}_{Rn}$ contains this error structure for the classical error as sub-dictionary. Moreover, the sub-dictionary contains the key 'proposal_sd' which denotes the initial standard deviation which is used to propose new values while the MCMC algorithm is in the sampling process. The sub-dictionary for a Berkson error only contains the information that the standard deviation is zero indicating no Berkson error. The sub-dictionaries 'exposure_model_distribution' and 'exposure_model_parameters' contain the distribution of $\mathcal{C}_{Rn}$ and its prior parameters respectively. The sub-dictionaries 'mapping_identifier_classical' and 'mapping_identifier_Berkson' specify whether the errors are shared between workers, within workers or both. As described before, the list ['year', 'object'], implies that all workers in M2 which share the same classical error when they are in a specific object and year. The key 'name_obs_values' identifies the error-prone variable in the data set. In the same way as for $\mathcal{C}_{Rn}$, the sub-dictionary for ω contains the uncertainty characteristics for this factor which can be read in the same manner. ω contains a Berkson error and therefore the standard deviation is now not set to zero and one has to specify the characteristics for the Berkson error as well. In the simulation study (Chapter 5), the Berkson error for ω is specified with a standard deviation of 0.69, is multiplicative and the initial standard deviation for the proposal is set to 0.1. We omitted the other factors for measurement model M2 for brevity. Note that for a standard deviation of 0.0 for the classical or Berkson measurement error, the algorithm does not account for the respective uncertainty in this part. If a user, for instance, would specify a classical and Berkson error with a standard deviation of 0.0, the algorithm would not correct for any measurement error in this factor and would only use the observed factor values (as provided in the data frame) as given values.

4.6.5 Structure of the algorithm

As the algorithm is implement in an object-oriented fashion, we are able to present the algorithm as a class diagram shown in Figure *4.4*. We refer to Appendix *A 1* for a description of the structure and class relations. The basic concept is to have a main class *MCMC* which controls the whole initialization, sampling procedure and saving of the results.

4.6.6 Documentation

We provide a documentation of the code as html-files. A file index.html serves as landing page and provides links to the documentation of the different python-modules.
We use *Sphinx* (Brandl, 2021) with *autodoc* as documentation tool. The documentation is therefore written at the head of each function or class as docstring in reStructuredText-format. When using *Sphinx* with *autodoc*, it is possible to extract the docstrings into .rst-files and afterwards to compile them to html. The theme and page layout stems from *Read the Docs*. More details can be found in the footer of the documentation.

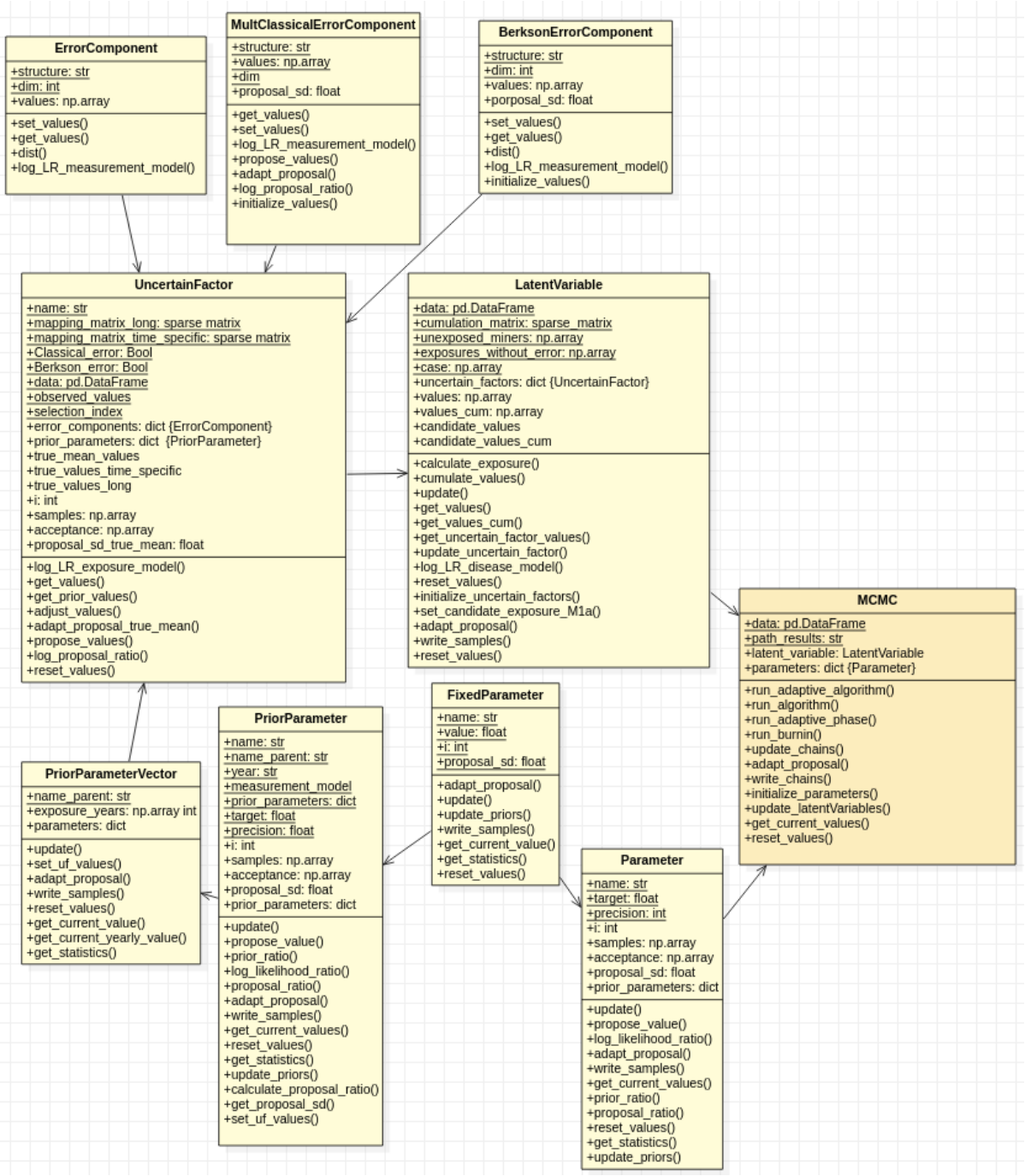

Figure 4.4: Class diagram for the implemented algorithm.

5 Simulation study

Before using the Bayesian hierarchical approach developed in Chapter *4* to account for measurement error on the real Wismut cohort data, we conduct a simulation study to assess and to compare the performance of the proposed Bayesian hierarchical approach with classical methods to account for measurement error and to study the impact of varying the complexity of the assumed measurement models.

In this chapter, we first give general information about the principles of simulation studies, then detail the aims of our simulation study and describe the design including the different simulation scenarios. In Section *5.4* we first describe the challenges we are faced with when generating exposure and survival data and then explain the generation of error-prone exposure data as well as the generation of survival times as a function of time varying exposure. After showing the used prior parameters for applying the Bayesian hierarchical approach, we describe SIMEX and regression calibration as two alternative methods to account for measurement error in Section *5.6* and present the results of the simulation study in Section 5.7.

5.1 General principles of simulation studies

When developing statistical methods to conduct parameter inference, it is important to assess the extent to which the proposed method fulfills some desirable statistical properties. In particular, a statistical method should produce parameter estimates $\hat{\theta}$ for an unknown parameter θ that are unbiased, i.e. the expected value of the parameter estimate $\hat{\theta}$ should be equal to the true parameter value θ or in other words, the parameter estimate should not suffer from systematic over- or underestimation. Similarly, it is important that the uncertainty intervals (i.e. confidence intervals in the case of frequentist estimators and credible intervals in the case of Bayesian estimators) are well calibrated in the sense that the probability to cover the true parameter value θ is close to the nominal coverage probability: When repeating an estimation procedure on 100 independent data sets to produce 95% confidence intervals, approximately 95 of these intervals should cover the true parameter value θ. In the case where more intervals cover the true parameter value, the interval overestimates the uncertainty associated with the estimation and the interval can be considered to be too conservative whereas in the case where less intervals cover the true parameter, the interval underestimates this uncertainty and it can be seen as too permissive.
While it is sometimes possible to derive theoretical results on the statistical properties of the proposed estimators, in particular in simple cases such as in the estimation of the mean of a sample or in the linear regression model, in more complex cases these properties are very difficult or even impossible to prove theoretically. Moreover, theoretical results often apply in asymptotic settings when the sample size goes to infinity. To assess the bias and the coverage probabilities of a proposed inferential procedure on finite samples, it is common to conduct simulation studies in which a large number of "fictive" data sets are generated according to a probability model with known and fixed parameter value θ to obtain parameter estimates $\hat{\theta}$ and confidence (or credible) intervals. By comparing the parameter estimates with the "true" parameter value θ which was used to generate the data sets, we can empirically assess the bias and coverage rates of the proposed statistical method. As it is possible to apply several estimation procedures on the same data set, simulation studies can also be used to compare the performance of competing methods.
When interpreting the results of simulation studies, one has to keep in mind that these studies can only give indirect evidence on the performance of statistical methods on real data sets: In a simulation study, "fictive" data sets are generated according to a probability model and it is a general truism that "all models are wrong" (Box, 1976). In this vein, a simulation study can only provide evidence on how a proposed statistical method would perform in a simplified world in which all (or at least a large part of) our modeling assumptions are fulfilled. To maximize the transferability of the results of simulation studies to real settings, it is desirable to generate "fictive" data sets that are as

close as possible to the real data that we are interested in. Moreover, simulation studies can be used to study the impact of model misspecification by applying an estimation procedure which assumes a simple probability model on data which was generated according to a more complex probability model. Finally, it should be mentioned that concepts like the bias of estimators or coverage rates are desirable properties of frequentist estimators and that their applicability to Bayesian methods of inference is somewhat controversial.

5.2 Aims of the simulation study

Besides investigating the applicability and the practical relevance of the Bayesian hierarchical approach (see Chapter *4*) to account for measurement error in the Wismut cohort, we pursue two aims with the simulation study.
The first aim of the simulation study is to compare the performance of our proposed method to account for exposure measurement error in the Wismut cohort with alternative approaches. Thus, we will compare our Bayesian hierarchical approach described in Chapter *4* with classical methods of correcting for measurement error, namely with SIMEX (see Section 5.6.1) and regression calibration (see Section 5.6.2).
The second aim of the simulations is to assess to what extent the complex structures of measurement error encountered in the Wismut cohort can be accounted for with measurement models which assume a simplified error structure. To address this point, we will generate exposure data according to a complex measurement model involving several types of measurement error (i.e. generalization error, parameter uncertainties etc.) and assume a more simple and therefore misspecified, measurement model when correcting for measurement error. By increasing the complexity of the assumed measurement models, it is possible to evaluate the adequacy of results of previous studies which aimed at correcting for uncertainty in radon exposure in cohorts of uranium miners. Additionally, we can thereby provide evidence to judge whether it is worthwhile to perform the extensive characterization and quantification of measurement error in an occupational cohort study as was performed by (Küchenhoff et al., 2018) and in Chapters *2* and *3*. Finally, this can also give evidence of whether it is worthwhile to develop a Bayesian hierarchical approach that is flexible enough to account for complex error structures with shared and unshared Berkson and classical components or whether it is also valid to use more classical methods of measurement error correction, such as regression calibration and SIMEX, which are easier to implement and which make a number of simplifying assumptions.

5.3 Design of the simulation study

In the design of the simulation study, our aim is to maximise the information gained through the simulations while respecting the limited time to compute all simulation scenarios as both the Bayesian hierarchical approach and SIMEX are very computationally expensive. In this vein, we focus on measurement model M2 (see Section *2.3*) and only use a Cox proportional hazards model with only one value for the true β in the simulations and a measurement error size quantified in Section *3.4*. We focus on M2 for the simulations since underground mining objects in the second and third exposure assessment period represent the largest proportion of person work years of all object types in the Wismut (together they represent 43.06% as can be seen in Figure *2.1*). The measurement models for underground mining objects in the second (M2) and third exposure assessment period (M3) only differ in one uncertain parameter: M3 involves the ventilation correction factor $c(o)$ instead of the equilibrium factor $g(t, o)$ which is present in M2. It would therefore be redundant to distinguish these two models in the simulation study and we choose M2 as it can be seen as more representative for the entire employment period of the Wismut cohort as it is between the first and the third exposure assessment period.

For the simulation study we consider three data generation scenarios (see Figure *5.1*) for which we generate exposure data with varying complexity. To assess the performance of our proposed

Bayesian hierarchical approach and to compare it with alternative methods to account for measurement error (see first aim of our simulation study), we estimate the parameter values assuming the respective measurement model, i.e. we do not consider model misspecification. In the comparison of the Bayesian hierarchical approach with SIMEX and regression calibration, we have to vary the complexity of the data generation process for the exposure data as it is not straightforward to assume mixtures of Berkson and classical error when accounting for measurement error via SIMEX.

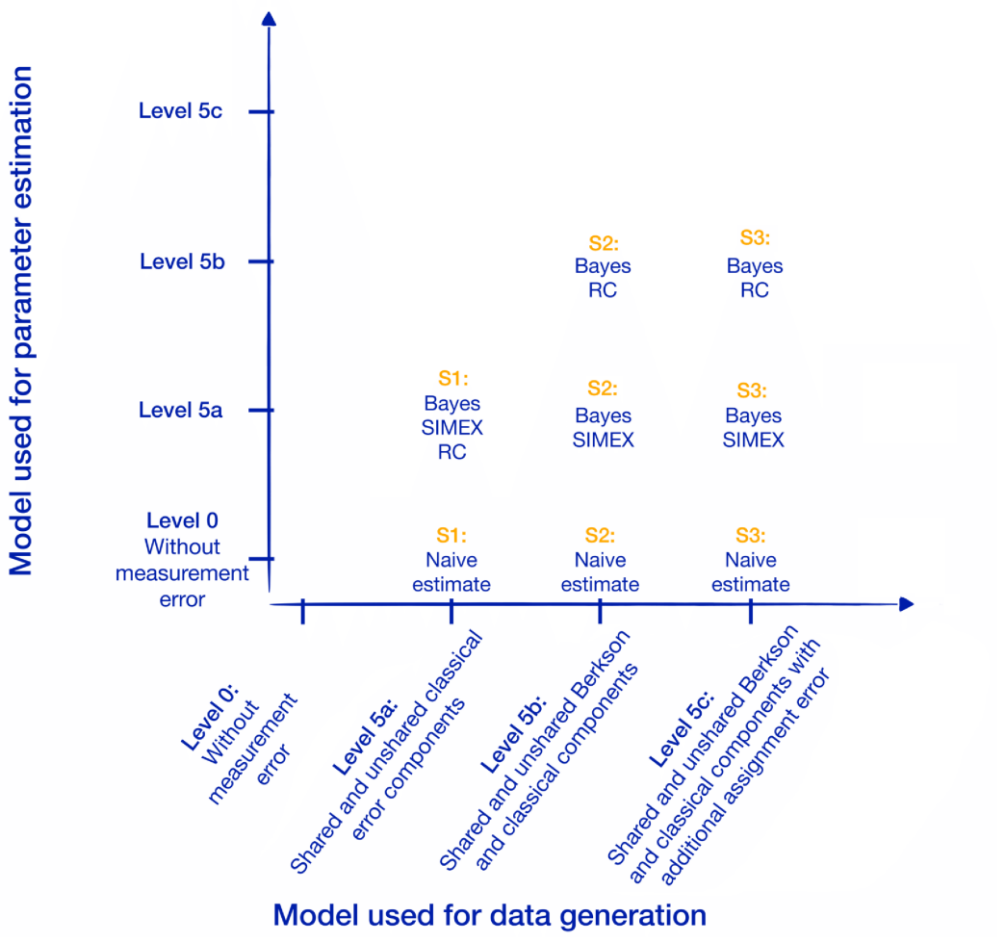

Figure 5.1: Overview of the three scenarios in the simulation study and the complexity levels of parameter estimation.

In this vein, we generate data according to complexity level 5 (see Table 4.1 in Section *4.1*). For simulation scenario 1 (S1) we generate data according to M2 with the exception that we neglect all Berkson error components (denoted as level 5a in Figure *5.1*) in order to be able to apply SIMEX besides the Bayesian hierarchical approach and regression calibration. In simulation scenario 2 (S2), level 5b in Figure *5.1*, we generate data according to M2 (Section *2.3*) without the Berkson assignment error and use the Bayesian hierarchical approach and regression calibration. In Figure *5.1*, these cases for scenarios S1 and S2 in which the model used for data generation is equivalent to the model used for parameter estimation are shown on the diagonal. As SIMEX is not able to account for Berkson uncertainties, in simulation scenario S2, in which exposure data is generated according to level 5b, we additionally correct parameter estimates for measurement error through SIMEX when neglecting all Berkson errors (i.e. we use level 5a for parameter estimation).

Additionally, to assess to what extent the complex structures of measurement error encountered in the Wismut cohort can be accounted for with measurement models which assume a simplified error structure (see second aim of our simulation study), we consider the results under model misspecification. In this vein we also study in simulation scenario 3 (S3) whether our decision to neglect assignment error when correcting for measurement error in the Wismut cohort is justifiable by first generating exposure data according to M2 on level 5 in Section *4.1* to then add an assignment error (i.e. we use level 5c in Figure *5.1* for data generation). We assume the assignment error to be additive and following Hoffmann et al. (2018b) to consist of one unshared Berkson error component reflecting the individual exposures of miners and years to vary around the group-based exposure and one Berkson error component that is shared for all years of exposure of the same miner, thus reflecting individual practices or job conditions of the miners. We assess the effect of neglecting assignment error by conducting parameter inference through the Bayesian hierarchical approach and regression calibration on level 5 (denoted as level 5b in Figure *5.1*) and through SIMEX on level 5 with the exception that all Berkson uncertainties are neglected (denoted as level 5a in Figure *5.1*). Furthermore, to assess the effect of model misspecification, we use the Bayesian hierarchical approach when assuming level 5a for parameter estimation (i.e. measurement error according to M2 without the Berkson parameter uncertainties) for data generated according to a more complex model including Berkson error components (scenarios S2 and S3). In Figure *5.1* the cases under the diagonal are those with misspecification.
In addition to this misspecification of the model structure, for scenario 3 we also consider misspecification of the size of the measurement errors by correcting for measurement error with the Bayesian hierarchical approach when using standard errors that are either half or double as large as the standard errors that were used to generate the data.
Finally, we produce a naive estimate which does not assume any measurement error in all three simulation scenarios to assess the overall impact of measurement error on risk estimation. We did not consider level 3 and level 4 neither as model to generate the data nor used for parameter estimation. Level 3 did not seem relevant for the Wismut since there is more detailed information on the different uncertain factors intervening in exposure estimation. We did not generate data according to level 4 because the only difference to level 5 is that all error components are assumed to be unshared and this does not seem to be realistic in the light of the exposure assessment in the Wismut cohort.

For each of the three simulation scenarios, we generate 100 data sets and estimate and compare a naive ($\hat{\beta}_{naive}$) and up to three corrected risk estimates ($\hat{\beta}_{Bayes}, \hat{\beta}_{RC}$ and $\hat{\beta}_{SIMEX}$) in order to derive the mean bias $\hat{\beta} - \beta$ and the mean relative bias $\frac{\hat{\beta}-\beta}{\beta}$ for the different estimators where β is the true parameter value which was used to generate the data. Moreover, we calculate coverage rates, i.e. the percentage of data sets for which the estimated 95% confidence or credible intervals cover the true parameter value β. Finally, we also consider the estimates of the baseline hazard parameters $\lambda_1, \lambda_2, \lambda_3$ and λ_4 to assess the impact of measurement error on the estimated piecewise constant baseline hazard.

5.4 Data generation

In this section, we describe how we generate the data for the simulation study on basis of a fictive data set of the Wismut cohort that we received from the Federal Office for Radiation Protection (BfS). First, we discuss the challenges when generating exposure and survival data for the simulations and then detail the generation of error-prone exposure data according to the measurement models of the considered simulation scenarios. In the third subsection we describe the generation of survival times with the annual radon exposure as time-varying covariate and time until death by lung cancer as censored outcome variable.

5.4.1 Challenges in the generation of exposure and survival data

As described in Section *5.1*, it is advisable to generate data sets that are as close as possible to the data set that we are interested in in order to maximize the transferability of the results in the simulation study. However, when generating exposure and survival data that reflect the characteristics of the Wismut cohort, we are faced with a number of challenges which we will discuss in this section.

Concerning the exposure data, it would be desirable to start from the exposure values observed in the Wismut cohort in order to be as close as possible to this data set. At the same time, we have to respect both the properties of classical measurement error - namely that the error terms have to be independent of the true exposure - and the properties of Berkson error - namely that the generated error terms have to be independent of the observed exposure. If we only had a classical measurement error affecting a miner's radon progeny exposure, we could assume the observed exposure values in the fictive cohort data set to be equal to the true exposure $X_i(t)$. To obtain the observed exposure $Z_i(t)$, we would need to either multiply or add an error term $U_i(t)$ which would be independent of true exposure and therefore respect the properties of classical measurement error: $Z_i(t) = X_i(t) \cdot U_i(t)$ or $Z_i(t) = X_i(t) + U_i(t)$. Conversely, if we only had a Berkson error on the radon progeny exposure, we could use the observed exposure values in the fictive cohort data set as observed exposure $Z_i(t)$ and either multiply or add a measurement error term $U_i(t)$ in order to obtain the true exposure either as $X_i(t) = Z_i(t) \cdot U_i(t)$ or $X_i(t) = Z_i(t) + U_i(t)$ (see for instance the simulation strategy of Küchenhoff et al. (2007) to generate Berkson error in the Wismut cohort). In contrast, a strategy that is not valid to generate Berkson error is to assume that the exposure data of the cohort are the true exposure values $X_i(t)$ and to generate observed exposure $Z_i(t)$ by dividing or substracting an error term $U_i(t)$: $X_i(t) = Z_i(t)/U_i(t)$ or $X_i(t) = Z_i(t) - U_i(t)$. As shown by Hoffmann et al. (2018a), this simulation strategy, which was used by Allodji et al. (2012b) and Allodji et al. (2012c) to study the impact of exposure measurement error when analyzing the association between radon exposure and lung cancer mortality in the French cohort and to assess the performance of two variants of regression calibration and of SIMEX in the same cohort, will erroneously produce systematic classical measurement error and therefore provide erroneous results.

Since the measurement models in the Wismut cohort, and in particular M2 (see Section *2.3*) involve several uncertain quantities, namely the radon gas concentration, the activity weighting factor, the working time factor and the equilibrium factor, with the latter three involving both Berkson and classical measurement error components, we need to generate exposure data according to a model in which we assume Berkson and classical measurement error components in several uncertain quantities simultaneously. Thus, it is simply not possible to assume either observed or true exposure in the simulation study to be the same as the observed exposure values in the cohort and one has to pay great caution in the generation of the exposure data in order to respect all underlying independence assumptions. The complexity of the measurement error model considered for the Wismut cohort therefore represents a major challenge for the generation of the simulation data.

In this situation, we have to first generate the true mean quantities (in the case of M2 (see Figure *2.7*) these quantities include the annual radon gas concentrations $\mathcal{C}_{Rn}(t, o)$, the activity weighting factor $\varphi(o, j)$, the equilibrium factor $\gamma(p_t, o)$ and the working time factor $\omega(p_t)$) according to so-called exposure models in which we make distributional assumptions and choose certain parameter values (see step 4 on page 87). Then, we add measurement errors which are generated according to measurement models for which we have to specify the variance parameters. Concerning the radon gas concentrations, we can for instance assume a normal or lognormal distribution with a certain mean and standard deviation. As the exposure models describe the distribution of latent variables, i.e. of quantities which are per definition not observable, it is difficult to make distributional assumptions and to find adequate parameter values for these variables.

Moreover, there are many sets of distributional assumptions and parameter values which yield

exposure data that is close to the observed exposure data in the cohort. To address this difficulty, we choose parameter values (for instance measurement error variance parameters) in accordance with the quantification of measurement errors for the Wismut cohort wherever this is possible and for the remaining parameter values which have to be specified (for instance the mean and the variance of the true and unknown radon gas concentration), we generate exposure data which is as close as possible to the fictive data set of the Wismut cohort provided by the BfS for the simulation study. The exact parameter values and the operationalization of this comparison will be described in Section *5.4.2*.

Concerning the survival data, the main challenge arises through the time-varying nature of exposure data in occupational cohorts and more specifically of radon exposure in the Wismut cohort. As described in Section *4.1*, we have to model radon exposure as a time-varying covariate if we want to describe measurement error on its natural level of occurrence (i.e. on the annual observed exposure values $Z_i(t)$ instead of on the cumulative radon exposure that a miner received during his entire employment period at the Wismut). However, when considering the cumulative radon exposure as time-varying covariate, the generation of survival times becomes very challenging since the standard approach of using the inversion of survival functions is not possible (Bender et al., 2005; Sylvestre and Abrahamowicz, 2008; Austin, 2012; Hendry, 2014). If the cumulative exposure would vary proportionally to time, one could use the method proposed by Austin (2012), which allows to simulate survival times with dichotomous time-varying covariates and continuous time-varying covariates $x(t) = kt$ which are proportional to time with a constant factor k. As the annual exposure in the Wismut cohort is not the same for every year, the cumulative exposure does not vary proportionally to time and this method is thus not possible for the Wismut. Alternatively, it is possible to use a permutation-based algorithm (Sylvestre and Abrahamowicz, 2008), but compared to the simple inversion of survival functions, this method is far less computationally efficient. We therefore adapt a method proposed by Zhou (2001) which is based on the generation of truncated piecewise exponential random variables to generate survival times as a function of continuous time-varying explanatory variables. This method was further extended by Hendry (2014) and Montez-Rath et al. (2017), who describe a number of pitfalls associated with this method which arise due to the sensitivity of the algorithm to some of the input parameter values which we will discuss in more detail in Section *5.4.3*.

5.4.2 Generating error-prone exposure data

Our aim for generating error-prone exposure data for the simulation study is to generate the measurement errors according to the measurement models considered in the simulation scenarios and to use as much information from the fictive data set of the Wismut cohort provided by the BfS as possible for the respective simulation scenario.

In this vein, we randomly choose $N = 1000$ miners from the fictive cohort data set for each simulated data set. To be in accordance with M2, the miners are chosen such that they worked at least one year in the respective exposure assessment period from 1955 to 1974. All of these miners' working years outside of this time period are cut off. As using the exposures from the fictive data set as either true or observed exposure data is not possible for simulation scenarios S1, S2 and S3, we pursue the goal to obtain error-prone observed exposure data $Z_i(t)$ that is as close as possible to the radon exposure values of the fictive cohort data set. Besides, we consider each exposure that is zero in the fictive data set to be known without error. Thus, we also set the simulated true and observed exposure ($X_i(t)$ and $Z_i(t)$) to zero for these years and miners.

Since the fictive cohort data set only contains the year of birth and the year of the miner's end of employment at the Wismut, we generate random dates containing an exact month and day for these two time points. Further, we assume that the annual radon exposures are received at the end of the respective years to the miners' working histories. Thus, the first radon exposure is received and

becomes effective on the 31st December of the year of start of employment at the Wismut. The only exception is the year of a miner's end of employment at the Wismut, where the exposure becomes effective on the exact day of end of employment. Besides, as the fictive cohort data set does neither contain any information about the objects the miners were working in nor the miners' activities, we choose arbitrary clusters of miners to represent the mining objects and the different activities. Therefor we assign each miner one of four objects and one of three activities, with one of the activities corresponding to the reference activity of a hewer. As in M2, we assume the value of the activity weighting factor for miners with the reference activity to be 1 and to be known without error.

To generate error-prone exposure data according to simulation scenarios S1, S2 and S3, for each underground mining object, we randomly choose from the respective distributions what we refer to as the true mean values of the uncertain parameters or the true mean radon gas or radon progeny concentrations. By multiplying the true mean values of the parameters or the true mean radon gas or radon progeny concentration with a classical error term, we obtain the observed mean values of the parameters or the observed mean radon gas or radon progeny concentrations, which were used in the radon exposure assessment. In case of an additional Berkson error component, the true mean values of the parameters are further multiplied with a Berkson error term in order to obtain the true values which may be specific to a year, an object and/or an activity. Using these true individual values of all uncertain parameters in the formula for the radon progeny assessment, we calculate the true radon exposure for each year $X_i(t)$. Finally, the error-prone observed exposure $Z_i(t)$ is calculated by using the observed values of all uncertain parameters in the formula for the radon progeny assessment.

In general, we need to specify a distribution for each unknown true mean value of the uncertain parameters $\omega(p_t)$, $\gamma(p_t, o)$ and $\varphi(o, j)$ as well as for the unknown true radon gas concentration $\mathcal{C}_{Rn}(t, o)$. As the ranges for the values of the uncertain parameters should be based on the Wismut cohort, we choose the true mean values from generalized beta distributions with shape parameters $\alpha = \beta = 3$ and with the respective support (e.g. the working time factor $\omega(p_t)$ is between 0.88 and 1.2 and the activity weighting factor $\varphi(o, j)$ lies between 0 and 1). In accordance to the Wismut cohort, we assume two periods for the equilibrium factor (until 1957, from 1958) and three periods for the working time factor (until 1958, 1959-1965, from 1966). For the true radon gas concentration, we use a truncated normal distribution with parameter values $\mu_{C_{Rn}}$ and $\sigma_{C_{Rn}}$ specified in such a way, that the distribution of the generated exposure data $Z_i(t)$ is as close as possible to the fictive data set. To determine this most suitable combination of parameter values, we choose all miners from the fictive cohort data set who worked at least one year during the second exposure assessment period from 1955 to 1974 and cut off the years before and after that period. We then generate the respective observed exposure values $Z_i(t)$ as described later, for a grid of different parameter values $\mu_{C_{Rn}}$ $(= 5{,}6, \ldots, 15)$ and $\sigma_{C_{Rn}}$ $(= 1{,}2, \ldots, 10)$ for the truncated normal distribution of the true and unknown radon gas concentration $\mathcal{C}_{Rn}(t, o)$. With the observed exposures $Z(t)$ for each combination of $\mu_{C_{Rn}}$ and $\sigma_{C_{Rn}}$, we calculate the respective sums of squared differences between the quantiles $(q_1 = 0.05, q_2 = 0.075, \ldots, q_{37} = 0.95)$ of the generated observed data $Z_i(t)$ and the subset of the fictive data that only contains miners who worked during 1955-1974 with years outside of this period cut off. To reduce the variability, we conduct this comparison for three different seeds and finally choose the combination of values for $\mu_{C_{Rn}}$ and $\sigma_{C_{Rn}}$ that leads to the lowest average sum of squared quantile differences.

Figure *5.2* shows for one simulation data set generated according to the three simulation scenarios S1, S2 and S3 the densities of the true and the observed annual radon exposure as well as the density of the respective exposures in the Wismut cohort data set. The densities shown are based only on the exposed miners, i.e. without non-exposed miners with 0 WLM. A direct comparison for time to death from lung cancer between the simulated data and the fictive data set of the Wismut cohort that we received from the BfS is not possible because the outcome variable was permuted in this

fictive data set. Therefore, Figure *A.2* in the Appendix *A 2* only shows the Kaplan Meier curves for non-exposed miners with 0 WLM and exposed miners for one simulated data set that was generated according to scenarios S1, S2 or S3.

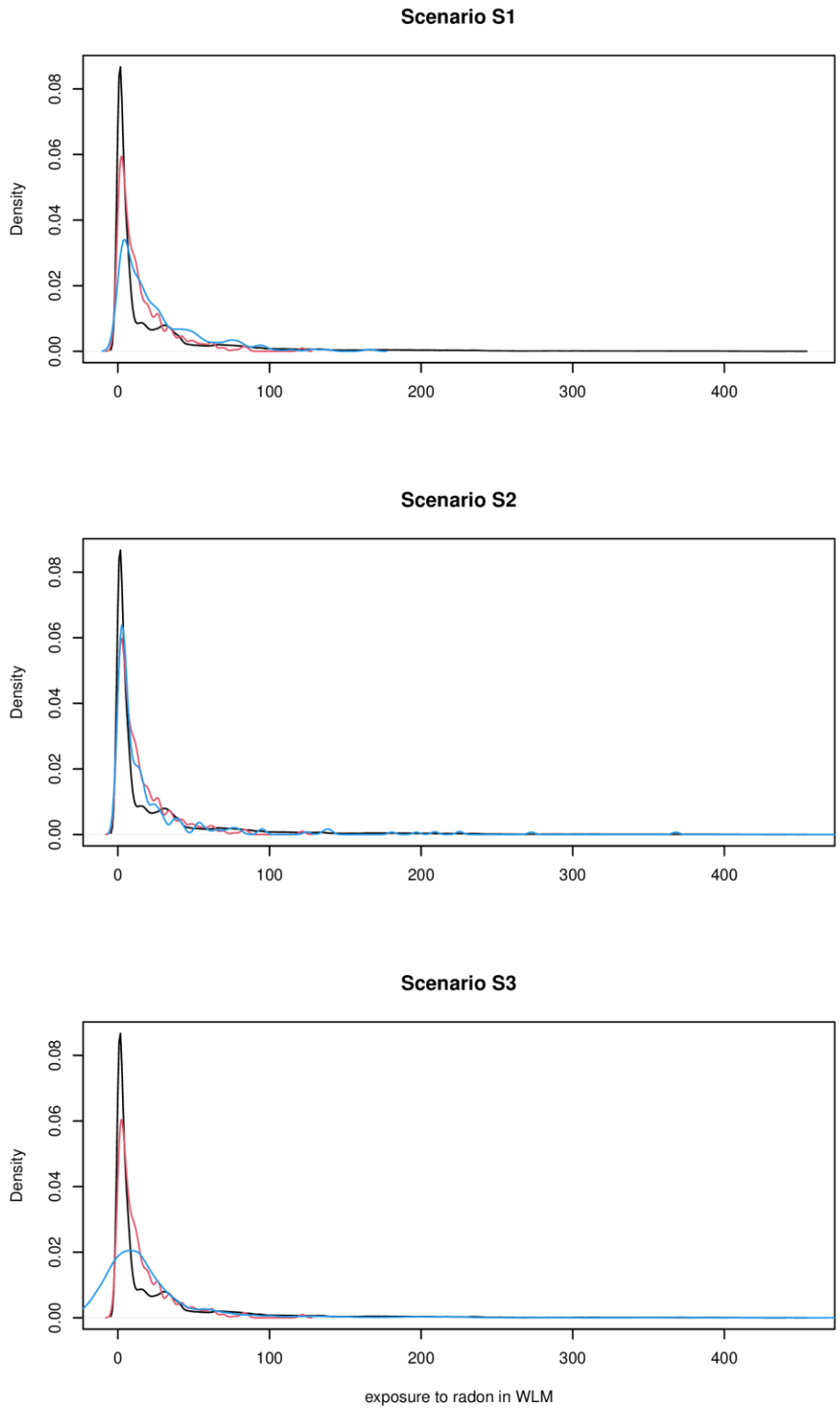

Figure 5.2: Densities of the true (blue) and the observed (red) radon exposure in WLM for one simulation data set generated according to scenarios S1, S2 or S3, respectively and the density of the exposures in the Wismut cohort data (black) between 1955 and 1974. Exposures of 0 WLM are not shown. Note that in scenario S3 some of the true radon exposures can be negative due to the additive assignment error.

In the following we summarize the steps to generate error-prone exposure data according to simulation scenario S2 (M2 including the Berkson error components for the equilibrium, working time and activity weighting factors but excluding the Berkson type assignment error).

1. Arbitrarily choose a day and month for a miner's birth and end of his employment at the Wismut.
2. Sample 1000 miners from those in the fictive data set, who worked at least one year in the time period from 1955 to 1974 and cut off their working time and exposure outside this period.
3. Choose arbitrary clusters of miners to represent four different objects and three different activities.
4. For each cluster of miners, sample the true mean values
   - Sample $\omega(p_t)$ from a generalized beta distribution with [0.88. 1.2] and shape parameters $\alpha = \beta = 3$
   - Sample $\gamma(p_t, o)$ from a generalized beta distribution with [0.2, 0.6] and shape parameters $\alpha = \beta = 3$
   - Sample $\varphi(o, j)$ from a beta distribution with [0,1] and shape parameters $\alpha = \beta = 3$
   - Sample $\mathcal{C}_{Rn}(t, o)$ from a normal distribution truncated at zero with parameter values $\mu_{C_{Rn}} = 6$ and $\sigma_{C_{Rn}} = 8$ which are determined through a comparison of the resulting generated values for $Z_i(t)$ (calculated as $Z_i(t) = C_{Rn}(t, o) \cdot 12 \cdot g(p_t, o) \cdot w(p_t) \cdot f(o, j)$ as described later) with the observed exposure values in the fictive data set.
5. Add to $\mathcal{C}_{Rn}(t, o)$ a classical error term $U_{\mathcal{C},c}(t, o)$ to obtain $C_{Rn}(t, o)$.
   $U_{\mathcal{C},c}(t, o)$ is sampled from a normal distribution with $\mu_{U_{\mathcal{C},c}} = 0$ and $\sigma_{U_{\mathcal{C},c}} = 0.59$ (as quantified for the generalization error in object 009 Aue in 1961).
6. Multiply the true mean value of the uncertain parameters with a classical error term
   - $f(o, j) = \varphi(o, j) \cdot U_{\varphi,c}(o, j)$
   - $w(p_t) = \omega(p_t) \cdot U_{\omega,c}(p_t)$
   - $g(p_t, o) = \gamma(p_t, o) \cdot U_{\gamma,c}(p_t, o)$

   where $U_{\varphi,c}(o, j), U_{\omega,c}(p_t)$ and $U_{\gamma,c}(p_t, o)$ are sampled from a lognormal distribution with $\sigma_{U_{\varphi,c}} = \sigma_{U_{\omega,c}} = \sigma_{U_{\gamma,c}} = 0.33$ (as quantified for the classical error component of the evaluation factor) and $\mu_{U_{\varphi,c}} = \mu_{U_{\omega,c}} = \mu_{U_{\gamma,c}} = -\frac{0.33^2}{2}$.
7. Multiply the true mean value of the uncertain parameters with a Berkson error term
   - $\varphi'(t, o, j) = \varphi(o, j) \cdot U_{\varphi',B}(t, o, j)$
   - $\omega'(t, o) = \omega(p_t) \cdot U_{\omega',B}(t, o)$
   - $\gamma'(t, o) = \gamma(p_t, o) \cdot U_{\gamma',B}(t, o)$

   where $U_{\varphi',B}(t, o, j), U_{\omega',B}(t, o)$ and $U_{\gamma',B}(t, o)$ are sampled from a lognormal distribution with $\sigma_{U_{\varphi',B}} = \sigma_{U_{\omega',B}} = \sigma_{U_{\gamma',B}} = 0.69$ (as quantified for the Berkson error component of the equilibrium factor) and $\mu_{U_{\varphi',B}} = \mu_{U_{\omega',B}} = \mu_{U_{\gamma',B}} = -\frac{0.69^2}{2}$.

8. Set $f(o,j)$ and $\varphi'(t,o,j)$ as well as $U_{\varphi,c}(o,j)$ and $U_{\varphi',B}(t,o,j)$ to 1 for miners with the reference activity since the activity weighting factor for a hewer is 1 and assumed to be known without error.
9. Calculate the true exposure value for miner i at time t $X_i(t)$ as $X_i(t) = \mathcal{C}_{Rn}(t,o) \cdot 12 \cdot \gamma'(t,o) \cdot \omega'(t,o) \cdot \varphi'(t,o,j)$.
10. Calculate the error-prone observed exposure for miner i at time t $Z_i(t)$ as $Z_i(t) = C_{Rn}(t,o) \cdot 12 \cdot g(p_t,o) \cdot w(p_t) \cdot f(o,j)$.
11. Set $X_i(t)$ and $Z_i(t)$ equal to zero for all years and miners who have an exposure of zero in the fictive data set.

The generation of exposure data for simulation scenario S1 is generally the same with the exception that all Berkson parameter uncertainties $U_{\varphi',B}(t,o,j)$, $U_{\omega',B}(t,o)$ and $U_{\gamma',B}(t,o)$ are set to 1.

For simulation scenario S3, we additionally add an additive assignment error, which consists of half an unshared and half a shared component by setting $X_i(t) = \mathcal{C}_{Rn}(t,o) \cdot 12 \cdot \gamma'(t,o) \cdot \omega'(t,o) \cdot \varphi'(t,o,j) + U_{E,B}(i,t,o,j) + U_{E,B}(i,o,j)$. The unshared error component $U_{E,B}(i,t,o,j)$ independently affects different miners as well as a miner's different exposure values, whereas the error component $U_{E,B}(i,o,j)$ is shared for all working years of the same miner in the same object and the same activity. In accordance with the quantification of the assignment error for object 009 (Aue) in 1961 as described in Section *3.4.3* and following Allodji et al. (2012a), we derive the standard deviation for both assignment error components as $\sigma_{U_{E,B}} = \sqrt{\frac{13.95^2}{2}} = 9.86$. Thus, $U_{E,B}(i,t,o,j)$ and $U_{E,B}(i,o,j)$ are sampled from a normal distribution with $\mu_{U_{E,B}} = 0$ and $\sigma_{U_{E,B}} = 9.86$.

5.4.3 Generating survival times as a function of time-varying exposure

For the simulation study we need to generate survival times with a time-varying covariate. The outcome of interest is the censored time until death by lung cancer (Y_i, δ_i), where δ_i describes the censoring indicator. The time-varying covariate is the sum of the annual radon exposure values received until time t, $X_i^{cum}(t)$.

In the following, we summarize the method proposed by Zhou (2001) which is based on the generation of truncated piecewise exponential random variables to generate survival times as a function of continuous time-varying explanatory variables as described by Hendry (2014) in short. The algorithm is based on generating a truncated piecewise exponential random variable V_i for each miner i. For that, a monotone increasing transformation function $g(.)$ has to be chosen in such a way, that $g(0) = 0$ and $g^{-1}(.)$ is differentiable. The baseline hazard rate of the Cox model according to which the survival times are generated, corresponds to $h_0(t) = \frac{\partial[g^{-1}(t)]}{\partial t}$. In order to incorporate the sum of the annual radon exposure values received until time t as a time-varying covariate, the rates $\lambda_i(t) = exp(\beta X_i^{cum}(t))$ have to be calculated for each miner i, where β is the regression coefficient. Besides, one has to define the bounds of truncation a and b, which correspond to the minimum and maximum possible generated time until death by lung cancer. The random variable V_i is drawn from a piecewise exponential distribution with varying rates $\lambda_i(t)$ and the corresponding time change points $g^{-1}(t)$. The truncation for the piecewise exponential variable is realized through an acceptation-rejection algorithm where only realizations within $g^{-1}(a)$ and $g^{-1}(b)$ are included. In the next step, the time until death by lung cancer of miner i is calculated as $T_i = g(V_i)$ and a censoring indicator $\delta_i \in \{0,1\}$ is defined (e.g. using an exponentially distributed censoring time), where for $\delta_i = 1$ miner i died of lung cancer, whereas he was censored for $\delta_i = 0$. Finally, all radon exposure values of miner i, which occurred after the censored survival time, are deleted.

One pitfall of the algorithm developed by Zhou (2001) and implemented by Hendry (2014) is the sensitivity of the distribution of the outcome to multiple user-supplied parameters. For this reason, Montez-Rath et al. (2017) performed simulation studies to evaluate these sensitivities and provide guidelines for the choice of the user inputs. Following these guidelines, the survival data should be generated using a large range for the bounds of truncation and the $g(.)$ function should be defined via a Weibull distribution with a respective Weibull shape parameter v.

To generate the time until death by lung cancer for the simulation study, we use the simulated true exposure values, cumulated for each year $X_i^{cum}(t)$ as time-varying covariate with a regression coefficient of $\beta = 0.003$ per WLM. Since we did not find a representative estimate of the risk estimate for the Wismut cohort in a Cox model in which radon exposure was treated as time-varying covariate, we used the risk estimate that was estimated in the French cohort of uranium miners. Moreover, as suggested by Montez-Rath et al. (2017), we use a wide limit for the generation of time until death by lung cancer for the algorithm by setting the lower bound of truncation for each individual miner to his age at begin of employment at the Wismut and the upper bound to 500 (since we truncate survival times at the age of 104, this upper bound of 500 years is somewhat arbitrary). Following the guidelines, we also use a Weibull distribution for the $g(.)$ function with a shape parameter of $v = 1$ to obtain a constant baseline hazard rate. However, we want to simulate according to a Cox proportional hazards model with a piecewise constant baseline hazard rate but this would violate the properties of the $g(.)$ function. We therefore set $h_0(t)$ to a constant value and integrate, depending on the age of the miner, the respective changing baseline hazard in the hazard rates $\lambda_i(t)$. In this way we get a piecewise baseline hazard that is constant in the four age ranges $[0,40)$, $[40,55)$, $[55,75)$ and $[75,104]$. These values were chosen in accordance with the patterns of lung cancer incidence in the general population. The upper limit of the last age interval results from the fact that we truncate the survival times at the age of 104, since this corresponds to the age of the oldest miner in the fictive cohort data set. Besides, we perform a censoring of the survival times by generating exponential distributed censoring times and defining the survival time as the minimum of the time until death by lung cancer and the censoring time. If a miner's censoring time is chosen smaller than his age at begin of employment at the Wismut, we draw a new value from the exponential distribution.

5.5 Applying the Bayesian hierarchical approach

We applied the Bayesian hierarchical approach as described in Chapter 4 to the data generated in the simulation study. Here, we show the used prior parameters in the simulation study. For the parameters of interest, we specify priors as follows:

- $\beta \sim N(0,100)$ (normal distribution parameterized with mean and standard deviation)
- $\lambda_1 \sim Ga(600,1/10000000)$ (Gamma distribution parameterized with shape and scale)
- $\lambda_2 \sim Ga(12000,1/1000000)$
- $\lambda_3 \sim Ga(46000,1/1000000)$
- $\lambda_4 \sim Ga(1000,1/100000)$

Parameters of the exposure models are specified as

- $\mathcal{C}_{Rn} \sim N^+(0.2,0.2)$ (normal distribution with values only on $\mathbb{R}^+$)
- $\varphi \sim B(3,3)$ (beta distribution with $a = b = 3$), truncated at $[0; 1]$
- $\omega \sim B(3,3)$ truncated at $[0.88,1.2]$
- $\gamma \sim B(3,3)$ truncated at $[0.2,0.6]$

5.6 Alternative methods to account for measurement error

As already described in the aims (Section *5.2*) and the design (Section *5.3*) of the simulation study, we use different methods for measurement error correction and risk estimation. While the Bayesian hierarchical approach is detailed in Chapter *4* and the naive risk estimate is simply obtained by using the observed exposure without any measurement error correction, in this section we describe how we performed the two alternative methods to account for measurement error, namely simulation extrapolation and regression calibration for the different simulation scenarios.

5.6.1 Simulation extrapolation (SIMEX)

Section *4.3.2* already briefly describes that using SIMEX for measurement error correction basically consists of the following three steps (see Carroll et al. (2006, Chapter 5) for more details):

1. Simulate data of increasingly larger measurement error $(1 + \lambda_{sim})\sigma_u^2$ with different values of $\lambda_{sim} \geq 0$ by calculating $Z_{b,i}(\lambda_{sim}) = Z_i + \sqrt{\lambda_{sim}}U_{b,i}$ for $i = 1, \dots, n$ and $b = 1, \dots, B$ and replacing X_i with $Z_{b,i}(\lambda_{sim})$ in the model calculation to finally obtain for each λ_{sim} value the average estimate $\hat{\beta}(\lambda_{sim})$ of the B risk estimates $\hat{\beta}_b(\lambda_{sim})$
2. Extrapolate back to $\lambda_{sim} = -1$ using an extrapolation function to obtain the SIMEX estimate
3. Calculate the standard errors

To account for measurement error in the Wismut cohort, SIMEX needs to be applied to survival data, which is possible by replacing X_i with $Z_{b,i}(\lambda_{sim})$ in the hazard function, and then using the partial likelihood to produce B estimates for each λ_{sim} (Carroll et al., 2006, pp. 323-324). With the average estimate $\hat{\beta}(\lambda_{sim})$ for each λ_{sim}, it is again possible to extrapolate back to $\lambda_{sim} = -1$ to obtain the SIMEX measurement error corrected estimate.

The SIMEX algorithm described so far corrects for additive classical measurement error, thus for the case that $Z_i = X_i + U_i$ with $U_i \sim N(0, \sigma_u^2)$. As we assume multiplicative errors for the equilibrium, working time and activity weighting factors in M2 (see Section *2.3*), we transform the multiplicative to the additive model $log(Z_i) = log(X_i) + U_i$, with $U_i \sim N(-\frac{\sigma_u^2}{2}, \sigma_u^2)$ to be able to apply SIMEX to multiplicative classical measurement error. Then, the simulation step is not performed on Z itself, but on its logarithm $log(Z)$, such that $Z_{b,i}(\lambda_{sim}) = exp(log(Z_i) + \sqrt{\lambda_{sim}}U_{b,i})$ (Carroll et al., 2006, pp. 104-108).

For the simulation study, we use SIMEX with the corresponding parameter values and a procedure adapted to simulation scenario S1 to S3 as described in the following.

1. Simulation step

   For each of the simulation scenarios S1 to S3 we use

   $$\begin{aligned} \lambda_{sim} &= \{0, 0.25, 0.5, 0.75, 1, 1.25, 1.5\}, \\ b &= \{1, 2, \dots, 100\}, \end{aligned}$$

   as well as the median as a more robust alternative to the mean to get the average of the B estimates for each λ_{sim}

   $$\hat{\beta}(\lambda_{sim}) = \text{median}(\hat{\beta}_b(\lambda_{sim})).$$

   For scenario S1, we assumed an additive classical measurement error for the radon gas concentration $C_{Rn}(t, o)$ and multiplicative classical errors for the three parameters $\omega(p_t), \gamma(p_t, o)$ and $\varphi(o, j)$. Scenario S2 additionally assumed Berkson error components for the working time, equilibrium and activity weighting factors and scenario S3 was equivalent to

scenario S2 with an additional Berkson assignment error. However, since SIMEX only corrects for classical error, we consider the following for all three scenarios S1, S2 and S3:

$$\begin{aligned} C_{Rn}(t,o) &= \mathcal{C}_{Rn}(t,o) + U_{\mathcal{C},c}(t,o) \text{ with } U_{\mathcal{C},c}(t,o) \sim N(\mu_{U_{\mathcal{C},c}}, \sigma^2_{U_{\mathcal{C},c}}) \\ log(f(o,j)) &= log(\varphi(o,j)) + U_{\varphi,c}(o,j) \text{ with } U_{\varphi,c}(o,j) \sim N(-\frac{\sigma^2_{U_\varphi}}{2}, \sigma^2_{U_\varphi}) \\ log(w(p_t)) &= log(\omega(p_t)) + U_{\omega,c}(p_t) \text{ with } U_{\omega,c}(p_t) \sim N(-\frac{\sigma^2_{U_\omega}}{2}, \sigma^2_{U_\omega}) \\ log(g(p_t,o)) &= log(\gamma(p_t,o)) + U_{\gamma,c}(p_t,o) \text{ with } U_{\gamma,c}(p_t,o) \sim N(-\frac{\sigma^2_{U_\gamma}}{2}, \sigma^2_{U_\gamma}) \end{aligned}$$

For each of the three parameters and the radon gas concentration we draw $U^{\varphi,c}_{b,i}(o,j), U^{\omega,c}_{b,i}(p_t), U^{\gamma,c}_{b,i}(p_t,o)$ and $U^{\mathcal{C},c}_{b,i}(t,o)$ from the respective normal distributions as described in the generation of error-prone exposure data in Section *5.4.2* and calculate

$$\begin{aligned} C_{Rn,b}(\lambda_{sim},t,o) &= C_{Rn}(t,o) + \sqrt{\lambda_{sim}} U^{\mathcal{C},c}_{b,i}(t,o) \\ f_b(\lambda_{sim},o,j) &= exp(log(f(o,j)) + \sqrt{\lambda_{sim}} U^{\varphi,c}_{b,i}(o,j)) \\ w_b(\lambda_{sim},p_t) &= exp(log(w(p_t)) + \sqrt{\lambda_{sim}} U^{\omega,c}_{b,i}(p_t)) \\ g_b(\lambda_{sim},p_t,o) &= exp(log(g(p_t,o)) + \sqrt{\lambda_{sim}} U^{\gamma,c}_{b,i}(p_t,o)) \end{aligned}$$

We replace the observed radon exposure with

$$Z_b(\lambda_{sim}) = 12 \cdot C_{Rn,b}(\lambda_{sim},t,o) \cdot f_b(\lambda_{sim},o,j) \cdot w_b(\lambda_{sim},p_t) \cdot g_b(\lambda_{sim},p_t,o)$$

and estimate the risk coefficients $\hat{\beta}_b(\lambda_{sim})$ to get

$$\hat{\beta}(\lambda_{sim}) = \text{median}(\hat{\beta}_b(\lambda_{sim})).$$

2. Extrapolation step

   Regardless of the underlying scenario, for each extrapolation step we use a simple quadratic extrapolation function

   $$\mathcal{G}_Q(\lambda_{sim}, (\alpha_1,\alpha_2,\alpha_3)^t) = \alpha_1 + \alpha_2\lambda_{sim} + \alpha_3\lambda^2_{sim}$$

   to extrapolate back to $\lambda_{sim} = -1$.

3. SIMEX standard error

   To get the SIMEX standard errors for each simulation scenario, we use the simulation-based method of nonparametric bootstrap. For this we generate 200 bootstrap samples, each containing 1000 miners randomly drawn with replacement from the respective simulation data set. For each of the 200 bootstrap samples, the SIMEX method is performed as described for the respective simulation scenario. To obtain the 95% credible interval, the fifth and the $195th$ of the ordered risk estimates of the 200 bootstrap samples is used as the lower and the upper limit respectively.

5.6.2 Regression calibration

As already mentioned in Section *4.3.1*, the regression calibration method has in general the following three steps (see Carroll et al. (2006, Chapter 4) for more details):

1. Estimation of the calibration function $E(X|Z)$ using some additional information like a calibration sample, repeated measurements or an instrumental variable. For the Wismut cohort, we can use the quantification for the measurement error magnitudes as this additional information.
2. Fit the standard model using $E(X|Z)$ instead of Z.

3. Calculate valid standard errors to account for the first step using bootstrap or sandwich methods.

As for SIMEX, in order to correct the data of the Wismut cohort for measurement error, we need regression calibration to apply to survival data and to additive as well as multiplicative measurement error. Regression calibration can be applied to survival data by simply using in the second step a standard model for survival analysis, e.g. the Cox model. Applying regression calibration to additive or multiplicative error differs in the first step, which is described in the following.

In the case of additive classical measurement error $Z_i = X_i + U_i$ with $U_i \sim N(0, \sigma_U^2)$ the calibration function $E(X_i|Z_i)$ can be estimated through

$$E(X_i|Z_i) = \mu_Z + \frac{\sigma_Z^2 - \sigma_U^2}{\sigma_Z^2}(Z_i - \mu_Z).$$

The multiplicative measurement error model can be transformed into the additive model $log(Z_i) = log(X_i) + U_i$ with $U_i \sim N(-\frac{\sigma_U^2}{2}, \sigma_U^2)$ and the calibration function is obtained with equation (4.7) in Carroll et al. (2006, p. 74) as

$$E(X_i|Z_i) = Z_i^{\frac{\sigma_{log(Z)}^2 - \sigma_U^2}{\sigma_{log(Z)}^2}} \; exp(\mu_{log(Z)}(1 - \frac{\sigma_{log(Z)}^2 - \sigma_U^2}{\sigma_{log(Z)}^2}) + \frac{\sigma_U^2}{2} + \frac{\sigma_{log(Z)}^2 - \sigma_U^2}{\sigma_{log(Z)}^2} \cdot \frac{\sigma_U^2}{2}).$$

The following describes how regression calibration is used in the simulation study for simulation scenarios S1 to S3.

1. Calibration function $E(X|Z)$

   Scenario S1:
   Simulation scenario S1 corresponds to M2 without any Berkson error components (level 5a in Figure *5.1*), so there is a classical additive measurement error for the radon gas concentration $\mathcal{C}_{Rn}(t, o)$ and only classical multiplicative errors for the activity weighting factor $\varphi(o, j)$, the working time factor $\omega(p_t)$ and the equilibrium factor $\gamma(p_t, o)$:

$$\begin{aligned}
C_{Rn}(t,o) &= \mathcal{C}_{Rn}(t,o) + U_{\mathcal{C},c}(t,o) \text{ with } U_{\mathcal{C},c}(t,o) \sim N(\mu_{U_{\mathcal{C},c}}, \sigma_{U_{\mathcal{C},c}}^2) \\
log(f(o,j)) &= log(\varphi(o,j)) + U_{\varphi,c}(o,j) \text{ with } U_{\varphi,c}(o,j) \sim N(-\frac{\sigma_{U_\varphi}^2}{2}, \sigma_{U_\varphi}^2) \\
log(w(p_t)) &= log(\omega(p_t)) + U_{\omega,c}(p_t) \text{ with } U_{\omega,c}(p_t) \sim N(-\frac{\sigma_{U_\omega}^2}{2}, \sigma_{U_\omega}^2) \\
log(g(p_t,o)) &= log(\gamma(p_t,o)) + U_{\gamma,c}(p_t,o) \text{ with } U_{\gamma,c}(p_t,o) \sim N(-\frac{\sigma_{U_\gamma}^2}{2}, \sigma_{U_\gamma}^2)
\end{aligned}$$

   To calculate the calibration function for scenario S1, we first use the formula for classical measurement errors for the radon gas concentration $\mathcal{C}_{Rn}(t, o)$ and the formula for multiplicative measurement errors for each of the three parameters $\varphi(o, j)$, $\omega(p_t)$ and $\gamma(p_t, o)$, thus

$$E(\mathcal{C}_{Rn}(t,o)|C_{Rn}(t,o)) = \mu_{C_{Rn}} + \frac{\sigma^2_{C_{Rn}} - \sigma^2_{U_{\mathcal{C},c}}}{\sigma^2_{C_{Rn}}}(C_{Rn}(t,o) - \mu_{C_{Rn}}),$$

$$E(\varphi(o,j)|f(o,j)) = f(o,j)^{\frac{\sigma^2_{log(f)} - \sigma^2_{U_\varphi}}{\sigma^2_{log(f)}}} \cdot exp(\mu_{log(f)}(1 - \frac{\sigma^2_{log(f)} - \sigma^2_{U_\varphi}}{\sigma^2_{log(f)}}) + \frac{\sigma^2_{U_\varphi}}{2} + \frac{\sigma^2_{log(f)} - \sigma^2_{U_\varphi}}{\sigma^2_{log(f)}} \cdot \frac{\sigma^2_{U_\varphi}}{2}),$$

$$E(\omega(p_t)|w(p_t)) = w(p_t)^{\frac{\sigma^2_{log(w)} - \sigma^2_{U_\omega}}{\sigma^2_{log(w)}}} \cdot exp(\mu_{log(w)}(1 - \frac{\sigma^2_{log(w)} - \sigma^2_{U_\omega}}{\sigma^2_{log(w)}}) + \frac{\sigma^2_{U_\omega}}{2} + \frac{\sigma^2_{log(w)} - \sigma^2_{U_\omega}}{\sigma^2_{log(w)}} \cdot \frac{\sigma^2_{U_\omega}}{2}),$$

$$E(\gamma(p_t,o)|g(p_t,o)) = g(p_t,o)^{\frac{\sigma^2_{log(g)} - \sigma^2_{U_\gamma}}{\sigma^2_{log(g)}}} \cdot exp(\mu_{log(g)}(1 - \frac{\sigma^2_{log(g)} - \sigma^2_{U_\gamma}}{\sigma^2_{log(g)}}) + \frac{\sigma^2_{U_\gamma}}{2} + \frac{\sigma^2_{log(g)} - \sigma^2_{U_\gamma}}{\sigma^2_{log(g)}} \cdot \frac{\sigma^2_{U_\gamma}}{2}),$$

where $\mu_{log(f)}, \mu_{log(w)}, \mu_{log(g)}, \sigma^2_{log(f)}, \sigma^2_{log(w)}$ and $\sigma^2_{log(g)}$ are the expectations and the variances of the logarithms of the values for $f(o,j), w(p_t)$ and $g(p_t,o)$, respectively. For the activity weighting factor, $\mu_{log(f)}$ and $\sigma^2_{log(f)}$ are calculated without those miners who belong to the reference activity of a hewer, since their activity weighting factor is assumed to be equal to 1 and to be known without error.

Finally, we simply combine the four calibration functions for the single measurement errors as follows to obtain the calibration function for M2 without the Berkson error components:

$$E(\mathcal{C}_{Rn}(t,o) \cdot 12 \cdot \gamma(p_t,o) \cdot \omega(p_t) \cdot \varphi(o,j)|C_{Rn}(t,o), g(p_t,o), w(p_t), f(o,j)) =$$
$$12 \cdot E(\mathcal{C}_{Rn}(t,o)|C_{Rn}(t,o)) \cdot E(\varphi(o,j)|f(o,j)) \cdot E(\omega(p_t)|w(p_t)) \cdot E(\gamma(p_t,o)|g(p_t,o))$$

Scenarios S2 and S3:

Compared to scenario S1, the scenarios S2 and S3 also consider Berkson error components for the activity weighting, the working time and the equilibrium factors and scenario S3 additionally considers the Berkson assignment error. As the assignment error is only considered in the model used for data generation and not in the model used for parameter estimation (see Figure *5.1*), regression calibration for S2 and S3 is the same. Calculating the single calibration function for the radon gas concentration $\mathcal{C}_{Rn}(t,o)$ is the same as in scenario S1, as there is no Berkson error component. Besides, when calculating the single calibration functions for each of the three parameters $\varphi(o,j)$, $\omega(p_t)$ and $\gamma(p_t,o)$, one can show that they are also the same as in scenario S1:

$$\begin{aligned}
E(\varphi'(t,o,j)|f(o,j)) &= E(\varphi(o,j) \cdot U_{\varphi',B}(t,o,j)|f(o,j)) \\
&= E(\varphi(o,j)|f(o,j)) \cdot E(U_{\varphi',B}(t,o,j)|f(o,j)) \\
&= E(\varphi(o,j)|f(o,j)) \\
E(\omega'(t,o)|w(p_t)) &= E(\omega(p_t) \cdot U_{\omega',B}(t,o)|w(p_t)) \\
&= E(\omega(p_t)|w(p_t)) \cdot E(U_{\omega',B}(t,o)|w(p_t)) \\
&= E(\omega(p_t)|w(p_t)) \\
E(\gamma'(t,o)|g(p_t,o)) &= E(\gamma(p_t,o) \cdot U_{\gamma',B}(t,o)|g(p_t,o)) \\
&= E(\gamma(p_t,o)|g(p_t,o)) \cdot E(U_{\gamma',B}(t,o)|g(p_t,o)) \\
&= E(\gamma(p_t,o)|g(p_t,o))
\end{aligned}$$

To obtain the calibration function for M2 including Berkson error, we again combine the single calibration functions as in scenario S1:

$$E(\mathcal{C}_{Rn}(t,o)\cdot 12\cdot\gamma'(t,o)\cdot\omega'(t,o)\cdot\varphi'(t,o,j)|C_{Rn}(t,o),g(p_t,o),w(p_t),f(o,j))=$$
$$12\cdot E(\mathcal{C}_{Rn}(t,o)|C_{Rn}(t,o))\cdot E(\varphi(o,j)|f(o,j))\cdot E(\omega(p_t)|w(p_t))\cdot E(\gamma(p_t,o)|g(p_t,o))$$

2. Standard model using $E(X|Z)$ instead of Z

   Instead of the observed radon exposure, we use the calibration function that was calculated in the first step for the respective simulation scenario and fit a Cox model with a piecewise baseline hazard that is constant in the four age ranges $[0,40)$, $[40,55)$, $[55,75)$ and $[75,104]$.

3. Regression calibration standard error

   We use nonparametric bootstrap to get the standard errors for regression calibration. For each simulation scenario we generate 200 bootstrap samples with 1000 miners drawn at random and with replacement from the respective simulation data set and then perform for each bootstrap sample regression calibration. The fifth and the $195th$ of the ordered risk estimates that are obtained for the 200 bootstrap samples are used for the lower and upper limit of the 95% credible interval.

5.7 Results of the simulation study

5.7.1 Results for β

Table *5.1* gives information on the performance of the proposed Bayesian hierarchical approach in comparison with regression calibration and SIMEX in simulation scenarios S1, S2 and S3 of the simulation study.

For simulation scenario S1, in which the data are only generated with shared classical error components in the four uncertain parameters ω, φ, γ and C_{Rn}, the naive estimates show a relative bias of about -10% and regression calibration, SIMEX and the hierarchical Bayesian approach are all able to reduce the absolute value of this relative bias to 5.96%, -4.24% and -2.95%, respectively. However, the coverage rate of the confidence intervals derived by the implemented version of regression calibration and SIMEX are 39% and 57% whereas the coverage rate of the credible intervals provided by the Bayesian hierarchical approach are 94% and thus very close to the nominal 95%.
For simulation scenario S2, in which the data are generated with shared Berkson and classical error components in the three uncertain parameters ω, φ and γ and with shared classical error for C_{Rn}, the naive estimates show a relative bias of close to 20%. Regression calibration and the proposed Bayesian hierarchical approach are again able to reduce the absolute value of this bias to -2.57% and 6.76%, respectively. Again, the Bayesian hierarchical approach provides credible intervals that result in a coverage rate of 93% which is very close to the nominal 95% whereas regression calibration provides only a coverage of 29%. When using SIMEX or a version of the Bayesian hierarchical approach that only accounts for classical error components (Level 5a in Figure *5.1*), the measurement model is misspecified. SIMEX and the Bayesian hierarchical approach produce coverage rates of around 60% and a bias of 4.88% and -11.47%, respectively, in this situation.
Finally, in simulation scenario S3, in which there are shared classical error components in the four uncertain parameters ω, φ, γ and C_{Rn} and shared Berkson error components for the factors ω, φ and γ as well as an assignment error in the form of a Berkson error consisting of half a shared and half an unshared component, strictly speaking all models are misspecified since none of the approaches accounts for this additional Berkson assignment error. In this scenario, the bias of the naive estimate is again around -20%. Both regression calibration and the Bayesian hierarchical approach can successfully reduce the absolute value of this relative bias to 3.50% and -3.91%, respectively. However, the coverage rates for the Bayesian hierarchical approach are again very close to the nominal level with 98% coverage whereas regression calibration only shows a coverage of 37%. The performance of SIMEX is comparable to its performance in simulation scenario S2 with a coverage rate of 60% and a relative bias of -13.81% and the version of the Bayesian hierarchical

approach that only accounts for classical error components (Level 5a in Figure *5.1*) leads to a small relative bias of 0.88% but the coverage rate is only 55%.
When the error variance parameters for the shared Berkson and classical error components in the four uncertain factors are misspecified by assuming either a standard deviation that is twice as big or half as big, the coverage rate of the Bayesian hierarchical approach is reduced to 28% and 80%, respectively. The relative bias, however is very high for the situation where the error variance is overestimated (178.8%) while it is negligible in the situation where the error variance is underestimated (-1.43%).
Figure *A.3* in the Appendix *A 3* shows for the Bayesian hierarchical approach for scenario S1 the estimated mean and credible interval for β on 100 simulated data sets. The horizontal dotted line is the true value of β.

Table 5.1: Results of the simulation study (rounded to two digits) for 100 data sets generated according to a scenario among S1 to S3 and with a true risk coefficient of β =0.3 per 100 WLM. As described in Section 4.1 and shown in Figure 5.1 'Level 5a' stands for adjustment only for classical errors according to M2 but without any Berkson error components. The naive frequentist and the naive Bayesian results are calculated by only using the disease model while assuming the observed exposure values to be without error.

	coverage rate	beta		bias of the mean	
		mean	median	absolute	relative in %
Scenario S1					
naive (frequentist)	0.31	0.27	0.25	-0.03	-11.32
naive (Bayes)	0.31	0.26	0.25	-0.04	-12.76
RC	0.39	0.32	0.27	0.02	5.96
Bayes	0.94	0.29	0.29	-0.01	-2.98
SIMEX	0.57	0.29	0.28	-0.01	-4.24
Scenario S2					
naive (frequentist)	0.25	0.25	0.24	-0.05	-17.36
naive (Bayes)	0.27	0.24	0.24	-0.06	-18.65
RC	0.29	0.29	0.25	-0.01	-2.57
Bayes	0.93	0.32	0.32	0.02	6.76
adjustment only for classical measurement error					
Bayes Level 5a	0.60	0.31	0.31	0.01	4.88
SIMEX	0.61	0.27	0.25	-0.03	-11.47
Scenario S3					
naive (frequentist)	0.28	0.24	0.24	-0.06	-19.27
naive (Bayes)	0.22	0.23	0.23	-0.06	-20.63
RC	0.37	0.29	0.25	-0.01	-3.91
Bayes	0.98	0.31	0.31	0.01	3.50
Bayes double size	0.28	0.84	0.80	0.54	178.76
Bayes half size	0.80	0.30	0.30	-0.00	-1.43
adjustment only for classical measurement error					
Bayes Level 5a	0.55	0.30	0.29	0.00	0.88
SIMEX	0.60	0.26	0.25	-0.04	-13.81

5.7.2 Results for the baseline hazard parameters

Table 5.2 shows results for the estimated baseline hazard parameters for the naive estimates, regression calibration and Bayes. For λ_2, λ_3 and λ_4, the proposed Bayesian approach shows good performance in correcting the baseline hazard parameter estimates. Depending on the scenarios and the parameter, the naive estimates can have a relative bias of more than 50%. While both error correction methods consistently produce overestimates of the true parameter values, the Bayesian hierarchical approach reduces these relative biases, whereas the regression calibration produces estimates that are comparable to the naive estimates. For λ_1, however, the Bayesian hierarchical approach shows estimates that are biased upwards by about 100% in all three scenarios. At the same time, the coverage rates for λ_1 are rather high (see Figure *A.4* in the Appendix *A 3*, which shows for the Bayesian hierarchical approach for scenario S1 the estimated mean and credible interval for the baseline hazards on 100 simulated data sets). The difficulty in estimating λ_1 can be explained by the fact, that there is very little information in the data to estimate the baseline hazard parameter for miners younger than 40 years of age. To palliate this problem, it would be advisable to use informative prior distributions for λ_1 on the real data of the Wismut cohort.

Table 5.2: Results of the simulation study for the estimates of the piecewise constant baseline hazard rate (rounded to two digits) for 100 data sets generated according to a scenario among S1 to S3 with the respective true values and the age intervals [0,40), [40, 55), [55, 75), [75, 104] for the baseline hazards $\lambda_1, \lambda_2, \lambda_3$ and λ_4.

baseline hazard parameter	λ_1			λ_2			λ_3			λ_4		
true value	$6.00 \cdot 10^{-5}$			$1.20 \cdot 10^{-3}$			$4.60 \cdot 10^{-3}$			$1.00 \cdot 10^{-2}$		
	mean in 10^{-5}	median in 10^{-5}	relative bias in %	mean in 10^{-3}	median in 10^{-3}	relative bias in %	mean in 10^{-3}	median in 10^{-3}	relative bias in %	mean in 10^{-2}	median in 10^{-2}	relative bias in %
Scenario S1												
naive	6.61	5.87	10.16	1.42	1.37	18.18	5.51	5.44	19.84	1.11	1.10	10.69
RC	6.56	5.96	9.36	1.43	1.40	18.98	5.51	5.42	19.83	1.09	1.07	9.26
Bayes	12.35	11.69	105.87	1.25	1.22	3.78	4.78	4.76	3.89	1.01	1.01	0.94
Scenario S2												
naive	8.74	6.74	45.58	1.84	1.71	53.57	5.77	5.71	25.36	1.08	1.06	8.11
RC	8.71	6.53	45.15	1.81	1.80	51.20	5.72	5.47	24.35	1.06	1.05	6.42
Bayes	11.83	11.52	97.08	1.27	1.22	5.81	4.75	4.75	3.29	1.02	1.01	1.81
Scenario S3												
naive	9.71	6.68	61.77	1.89	1.83	57.43	5.89	5.75	28.03	1.07	1.05	6.84
RC	9.54	6.62	58.96	1.86	1.84	55.16	5.81	5.71	26.21	1.05	1.03	5.46
Bayes	12.79	12.28	113.18	1.35	1.33	12.70	4.90	4.83	6.60	1.00	1.00	0.16

5.7.3 Conclusions on the practical relevance of the proposed method for the Wismut cohort

The results of the simulation study confirmed that exposure measurement error generally leads to an attenuation in risk estimates. Moreover, it showed that the proposed Bayesian hierarchical approach is able to correct for the complex error structure in the Wismut cohort. In contrast to regression calibration and SIMEX, this approach provides good coverage rates for all three simulation scenarios and seemed in particular advantageous when correcting for shared error components. Moreover, the comparison between the correctly specified Bayesian hierarchical approach and the misspecified one, which only corrects for classical error components, shows that it is important to consider the shared Berkson error components of the three factors ω, φ and γ when correcting for measurement error in the Wismut cohort. However, it does not seem to be necessary to account for the assignment error in the Wismut cohort. In simulation scenario S3, the Bayesian hierarchical approach shows a good overall performance in correcting for the bias introduced by exposure measurement error and it shows a good coverage rate, even though it neglects this Berkson error with an unshared and a shared component. The simulation study also underlined the importance of an extensive exposure quantification as it showed that misspecifications of the error variances can result in very high relative biases.

6 Accounting for exposure uncertainty in the Wismut cohort

6.1 Necessary adaptions to account for measurement error in the Wismut cohort

In order to account for measurement error in the Wismut cohort, we combined the proportional hazards model without effect modification described in Section *4.4.1* with the measurement models described in Chapter 2. In order to do so, the real cohort data had to be extended. The respective measurement model and the corresponding observed uncertain parameters had to be added to the Wismut data in long format, i.e. the data frame contains several rows per miner with one row for each time point at which the miner's radon exposure changed and at which the miner was either censored or died of lung cancer. For this purpose, the parameters required for the exposure assessment were determined from Lehmann et al. (1998) and Lehmann (2004) for all years of each object. In cases where the exposure assessment differed within one object, the required parameters were determined for each shaft group of the respective object. The actual creation of the extended cohort data with this information was mainly carried out by BfS.

The measurement models described in Chapter *2* according to Lehmann et al. (1998) and Küchenhoff et al. (2018) were in many cases applicable to the real data of the Wismut cohort. However, when accounting for the more detailed and shaft specific exposure estimation according to JEM 2 (Lehmann, 2004) and when trying to add the necessary information to the Wismut cohort data in order to obtain all parameter values and necessary information to account for these measurement models, we encountered a variety of examples where it was very difficult to apply these models exactly in the way that we had planned when we characterized and quantified exposure uncertainty. In such cases, we had to adapt the measurement error models derived in Chapter *2* or the dependence structures that were assumed in these models and therefore as well the quantified values of measurement error variances. This was necessary either because the exact information on the quantities was missing or because the exposure assessment was more complex than initially described in Lehmann et al. (1998). Even with these adapted measurement models it was unfeasible to model the exposure of miners who were listed under the object number 090 000 as this object number is not part of the JEMs (Lehmann et al., 1998; Lehmann, 2004). Since this only applied to a handful of miners, they were excluded from the data set. As we will describe in more detail in Section *6.1.3*, we assumed a measurement error model that was similar to M4 in cases in which we lacked information to reconstruct the detailed exposure values according to the initial measurement model or in which the exposure values in the extended cohort data could not be reproduced with the original models. In cases in which the annual exposure values derived based on Lehmann et al. (1998) or Lehmann (2004) only slightly deviated from the annual exposure values from the extended cohort data or in which it seemed more reasonable to keep the original measurement model, we did not modify the assumed measurement model, but defined a multiplicative term τ_{post} with which the exposure values would need to be multiplied to coincide with the exposure value from the extended cohort data. τ_{post} was calculated with the aim to assess the robustness of the results with respect to the multiplication of this term, but it was not in the scope of the present study to perform such a sensitivity analysis.
Finally, in line with the results of the simulation study, we decided that it is not worthwhile to account for assignment error in the Wismut cohort, so we adapted all considered measurement models by omitting the assignment error.

6.1.1 Adapting dependence structures between measurement error components

Adapting dependence structures for parameter uncertainties

It was originally assumed that the observed values for the different parameters would be different for different years, objects and activities. The observed value of the activity weighting factor $f(o, j)$, for instance, was assumed to depend on object o and activity j. However, in the real data these

values often coincided. If we had access to more detailed information on the exact object and activity, we could use more precise information provided in Lehmann et al. (1998) to decide whether these observed values just happen to coincide or were really estimated jointly. However, because of data protection rules, we do not actually have access to the exact object and activity of the miners and therefore cannot use the information provided by Lehmann et al. (1998) to define the dependence structures of the occurring classical error components. Even if we had this information, it would be difficult to decide whether two values that are exactly the same were indeed estimated independently. To answer this question, we would need to interview the experts who were involved in the exposure assessment of the Wismut cohort.
As a consequence, we will describe the set of objects, years and activities for which we assume shared classical errors as "clusters" in the following. For instance, we now assume $f(p_{oj})$ where p_{oj} can be thought of as the result of a function $p(o,j)$ that assigns the same value p_{oj} to all objects and activities for which the observed values of the activity weighting factor $f(o,j)$ are the same. Implicitly, this function $p(o,j)$ would in this example define clusters of objects and years for which the activity weighting factor is constant. While we originally mainly assumed such clusters $p_t = p(t)$ for the working time factor where these clusters merely reflected time periods, we now assume clusters that can also depend on objects and activities. For the working time factor $w(p_t)$ we assume a classical error component $U_{\omega,c}(p_t)$ which is shared for all years of one cluster with the same observed working time factor as this implies that the implicit estimation was done simultaneously for all years and objects for which this value is used. Similarly, we assume for the equilibrium factor, the ventilation factor, the activity weighting factor and the evaluation factors for open pit mining as well as for the transfer factor for the evaluation factor and the proportion of exposure from old mining that their respective classical error components are shared for all years and objects for which they have the same observed value. In the updated DAGs shown in the following, these decisions are reflected in the modification of the arguments that the uncertain parameters take.

Adapting measurement error characteristics to account for errors arising in the transfer of mean radon progeny and radon gas concentration measurements

Many of the adaptations of the JEM 2 consist in using the JEM 1 exposure value of a specific year (transfer year) and object (transfer object) to transfer these values to another year and/or object. This is very similar to the assessment of exposure values for development and exploration objects where the exposure values of an affiliating mining object are multiplied by a transfer factor τ_E as described in Chapter *2*. For example, in the case of a development object for which the exposure is 30% of the exposure from the affiliated mining object, the exposure from this affiliated mining object and the respective year is transferred to the development object and multiplied with $\tau_E = 0.3$. We assume that the development object has the same measurement model and the same uncertain exposure values as the transfer object in the transfer year, hence, in this case the same measurement model and the same uncertain exposure values as the affiliated mining object in the same year. In the same way, we treat cases for which the exposure assessment for a given object and year has been adjusted by the JEM 2 by using the exposure values of another object (transfer object) and year (transfer year) and by multiplying them with a transfer factor τ_{JEM}. Again, the given object and year is assumed to have the same measurement model and the same uncertain exposure values as the transfer object in the transfer year. In this vein, we assume for instance that for objects for which the exposure assessment for 1988 and 1989 was not explicitly described, but the same radon progeny exposure values as from the previous year 1987 were given in the JEM 2 tables, that this exposure value from 1987 was continued. Since we treat the two transfer factors τ_E and τ_{JEM} exactly in the same way, we will only speak of transfer factor $\tau = \tau_E \cdot \tau_{JEM}$ in the following.

Initially, we had planned to model the transfer of exposure values by assuming a classical and a Berkson error component for the uncertain parameter τ_E. However, when working with the extended Wismut cohort data, it became clear that it made more sense to model the transfer of

exposure values by adapting the dependence structures for the mean radon gas and radon progeny concentration measurements and by assuming a Berkson error component for these uncertain quantities. In this vein, we assume no error on the transfer factor τ, thus, multiplying with $\tau = 1$ has no effect on the exposure value itself. However, the transfer of exposure values from one year and object to another creates a dependence in the measurement error terms that we have to account for. Indeed, the classical error that occurs in the estimation of the exposure value for the object for which it was originally estimated is now shared between this original object and the object this exposure value is transferred to. It influences the cluster of observations for which we assume a shared error component for the mean radon gas or radon progeny concentration measurement and for the uncertain parameters that were used for this year and object. In addition to this shared classical error component for the radon gas or radon progeny concentration measurements, we account for the fact that the exposure value of the object and year that an exposure value was transferred to can differ from the estimate in the transfer year and/or transfer object. In this sense, if a cluster of observations consists of five non-transfer objects and years and one object and year that served as the transfer object and year, we assume a multiplicative object and year specific Berkson error for all five non-transfer objects. In cases in which there was a joint estimation without any clear reference object, we assumed a Berkson error for all obejcts and years in the cluster.

6.1.2 Adapting the dependence structure of Berkson errors to account for shaft specific exposure values

We slightly had to adapt the structure of the Berkson error for the equilibrium factor, because there were instances where different values of this factor were assumed for different shafts of the same object in the same year. As a consequence, we used a shaft specific Berkson error component for the equilibrium factor. For the same reasons, we now assume a shaft specific Berkson error for radon gas and radon progeny concentration in M2 and M3, respectively.

6.1.3 Adapting measurement models

The measurement models as described in Chapter *2* have to be adjusted because of the newly adapted dependence structures between error components, but also because of various problems that have arisen. For example, some information was not available that would have been needed for the originally assumed measurement model, or for some special cases the exposure assessment was carried out differently, resulting in new measurement models. Note also that the changes in the error structures for radon gas and radon progeny concentration measurements that were described in the last section made the measurement error that was originally assumed for τ redundant and we will therefore no longer assume an error for this quantity.

In the following we will briefly describe the adapted measurement models.

M0

We assumed measurement model M0 with an exposure to radon progeny of 0 WLM and no measurement error whenever at least one of the following applied

1. a miner worked in a surface object, even if the extended data of the Wismut cohort contained no value for the exposure (WLM was missing in the data)
2. the object and object type were unknown but radon exposure was specified as 0 WLM
3. the miner's activity weighting factor had a value of 0, which, when multiplied to a hewer's radon progeny exposure, also gives an exposure of 0 WLM.

M1a and M1b

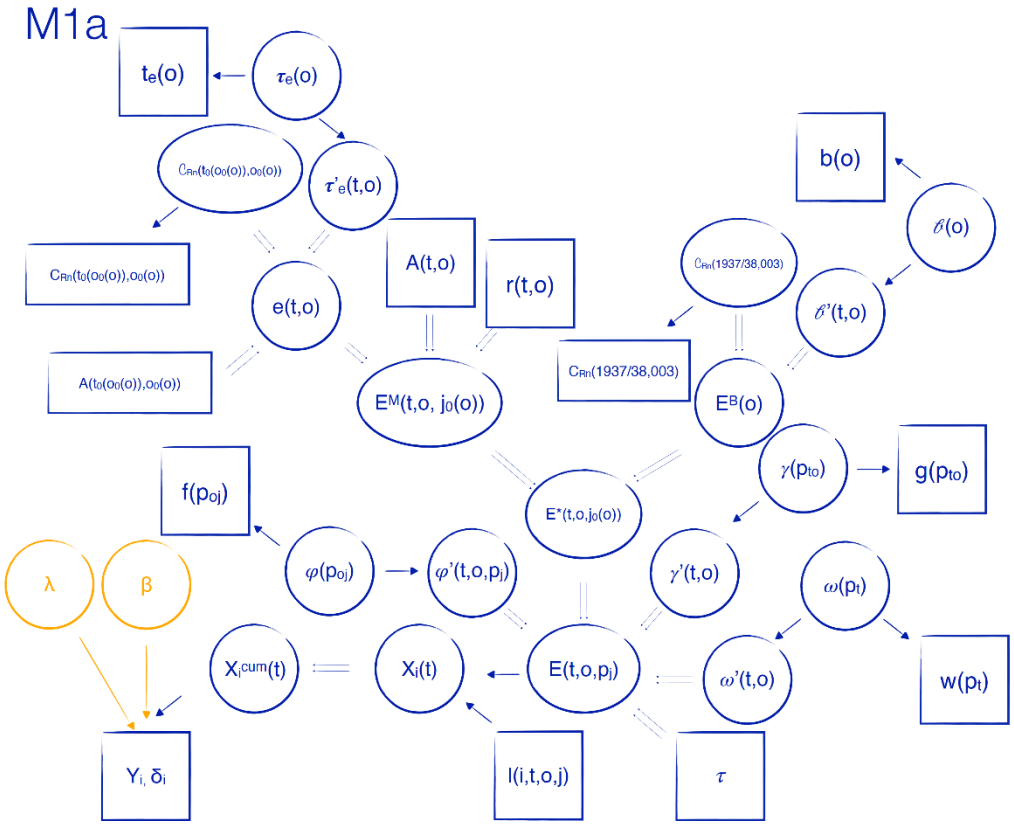

Figure 6.1: Hierarchical model combining a disease model with measurement model M1a to describe how exposure uncertainty in underground mining objects as well as in exploration and development objects in Saxony was accounted for on the data of the Wismut cohort in the first exposure assessment period. Due to the limited space and for a clearer presentation, no measurement error variances are shown here.

According to Section *2.2*, the evaluation area $A(t,o)$ for objects in Saxony during the first exposure assessment period (measurement model M1a) is calculated as $A(t,o) = C(t,o) + p\sum_{s=1946}^{t-1} C\,(s,o)$ with the mined vein area $C(t,o)$ and p the proportion of mined vein area from previous years $C(s,o)$. However, the size of the mined vein area or the cumulative mined vein area from previous years is unknown for several years and objects. For those objects and years for which the information on $C(t,o)$ and $\sum_{s=1946}^{t-1} C\,(s,o)$ is neither given in Lehmann et al. (1998) nor Lehmann (2004), this information is not available for this project. With missing information on the (cumulative) mined vein area, it is impossible to use measurement model M1a as described in Chapter *2* and we therefore had to adjust it in such a way that for these years and objects the evaluation area $A(t,o)$ was calculated from the radon exposure from the Wismut cohort members $Z_i(t,o)$ and the known values of the remaining parameters.

$$A(t,o) = \left(\frac{Z_i(t,o)}{g(p_t,o)\cdot w(p_t)\cdot f(o,j)\cdot 12} - C_{Rn}(1937/1938{,}003)\cdot b(o)\right)\cdot \frac{1}{r(t,o)\cdot t_e(o)}\cdot\frac{A(t_0(o_0(o)),o_0(o))}{C_{Rn}(t_0(o_0(o)),o_0(o))}$$

When correcting for measurement error according to measurement model M1a we now consider the evaluation area $A(t,o)$ to be known without measurement error. Since for measurement model M1b this assumption of the evaluation area $A(t,o)$ to be known without measurement error was already made in Chapter *2*, the measurement model for M1b does not change here.
In situations where the radon exposure estimation according to Lehmann et al. (1998) or Lehmann (2004) was solely based on the radon exposure from old mining without any exposure from mining

activity, measurement model M1a was applied in a reduced form. Only the uncertain parameters for old mining $C_{Rn}(1937/1938{,}003)$ and $b(o)$ as well as the equilibrium, working time and activity weighting factors with the respective classical and Berkson measurement error components were considered. The measurement model in this reduced form corresponds to the DAG shown in Figure *6.1*, except that the exposure from mining activity $E^M(t, o, j_0(o))$ and everything pointing there is not to be taken into account.

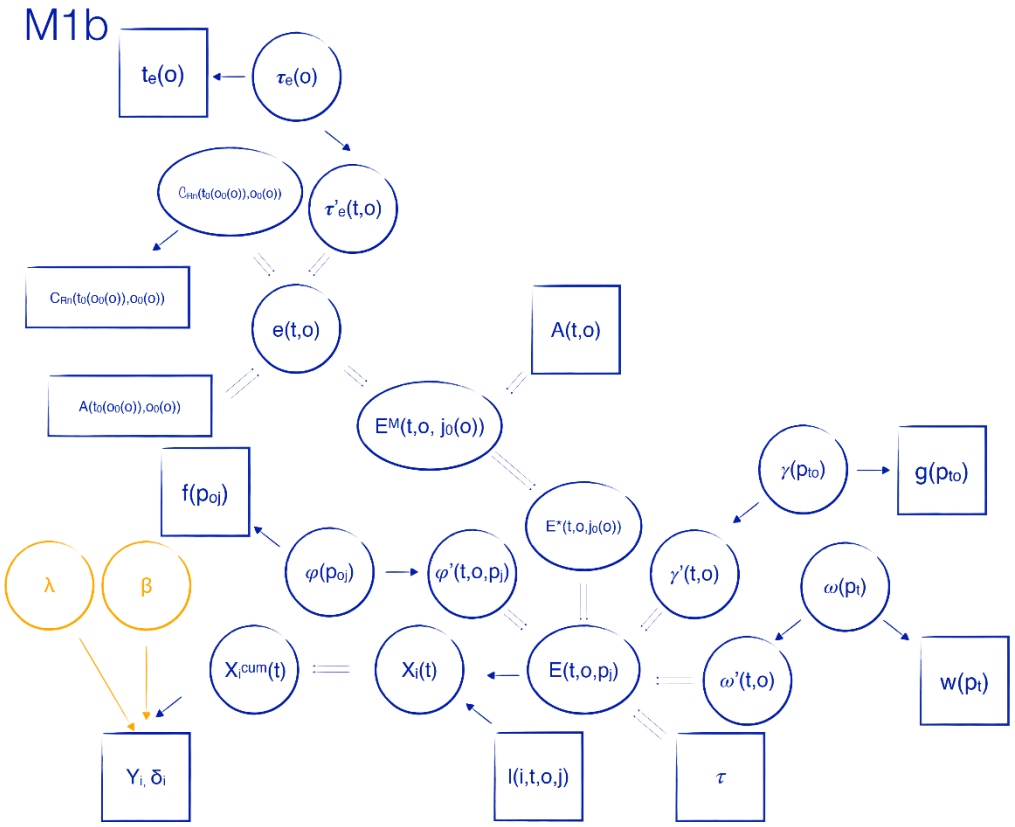

Figure 6.2: Hierarchical model combining a disease model with measurement model M1b to describe how exposure uncertainty in underground mining objects as well as in development objects in Thuringia was accounted for on the data of the Wismut cohort in the first exposure assessment period. Due to the limited space and for a clearer presentation, no measurement error variances are shown here.

For cases in which the exposure of an object and year according to measurement models M1a and M1b was transferred to another object and/or year, this object and/or year gets the same measurement model with the corresponding parameter values and the additional factor τ. For M1a we assume that all old mining objects belong to one cluster for which the classical error of the radon gas exposure measurements from old mining $C_{Rn}(1937/1938{,}003)$ are shared. Besides, each old mining object belongs to one cluster depending on the value of the proportion from old mining $b(o)$. For all objects in one cluster, thus, with the same observed $b(o)$ value, a shared classical error component $U_{b,c}(o)$ is assumed for $b(o)$ (see also Section *6.1.1*), whereas the Berkson error $U_{b,B}(t, o)$ is assumed as unshared for all objects except for the reference object 003 Schneeberg, where the old mining measurements $C_{Rn}(1937/1938{,}003)$ were taken.
In the same way we also assume that each object with exposure assessment according to measurement models M1a or M1b belongs to one of four clusters, for which the classical error for the radon gas concentration measurements $U_{C,c}(t_0(o_0(o)), o_0(o))$ is assumed as shared depending on which of the four reference objects was used for $C_{Rn}(t_0(o_0(o)), o_0(o))$. All objects with the same value of the transfer factor for evaluation $t_e(o)$ belong to one cluster with shared classical error component $U_{\tau_e,c}(o)$ (see also Section *6.1.1*). Besides, an unshared Berkson error component $U_{\tau_{e'},B}(t, o)$ is assumed for all years of all objects except the four reference objects for which there is

no transfer factor for evaluation t_e as the evaluation factor $e(t,o)$ was calculated for these objects and not transferred.

Figures *6.1* and *6.2* show the adapted DAGs for objects in Saxony and Thuringia, respectively, for the first exposure assessment period in the Wismut cohort.

M2 and M3

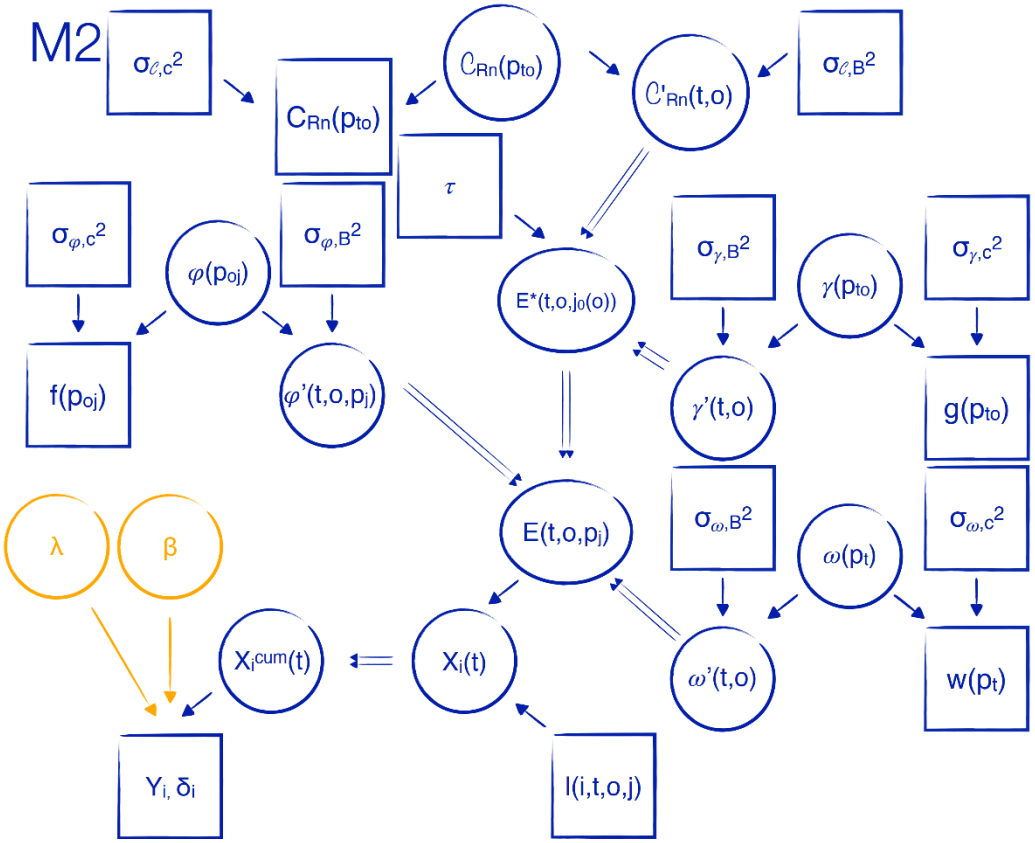

Figure 6.3: Hierarchical model combining a disease model with measurement model M2 to describe how exposure uncertainty in underground mining objects in Saxony and Thuringia as well as in development objects was accounted for on the data of the Wismut cohort for the second exposure assessment period.

The measurement error model for the second exposure assessment period (M2) as derived in Chapter *2* assumed an additive classical generalization error on the mean annual radon gas concentration C_{Rn} since the averaging of several measured radon gas concentration values can be assumed to be normally distributed. For individual years of some objects, however, the radon exposure estimation was not based on averaged measured concentration values, but the radon gas concentration was determined by experts. This was the case when there were no measured values for that year, or when the measurements appeared implausible and were corrected accordingly by experts. For such years and objects, we adopt a new measurement error model M2_Expert, which basically corresponds to the exposure assessment of M2, but assumes a multiplicative classical expert estimation error $U_{\mathcal{C}_{Exp},c}(p_{to})$ on the annual radon gas concentration instead of an additive classical generalization error $U_{\mathcal{C},c}(p_{to})$. Assuming a multiplicative error guarantees the positivity of both true and observed exposure. Moreover, assuming lognormal measurement errors is in line with a large part of the literature on exposure uncertainty in radon exposure (Lubin et al., 1995b; Stram et al., 1999; Heid, 2002; Heid et al., 2004; Lubin et al., 2005; Heidenreich et al., 2012; Allodji et al., 2012a,b,c).

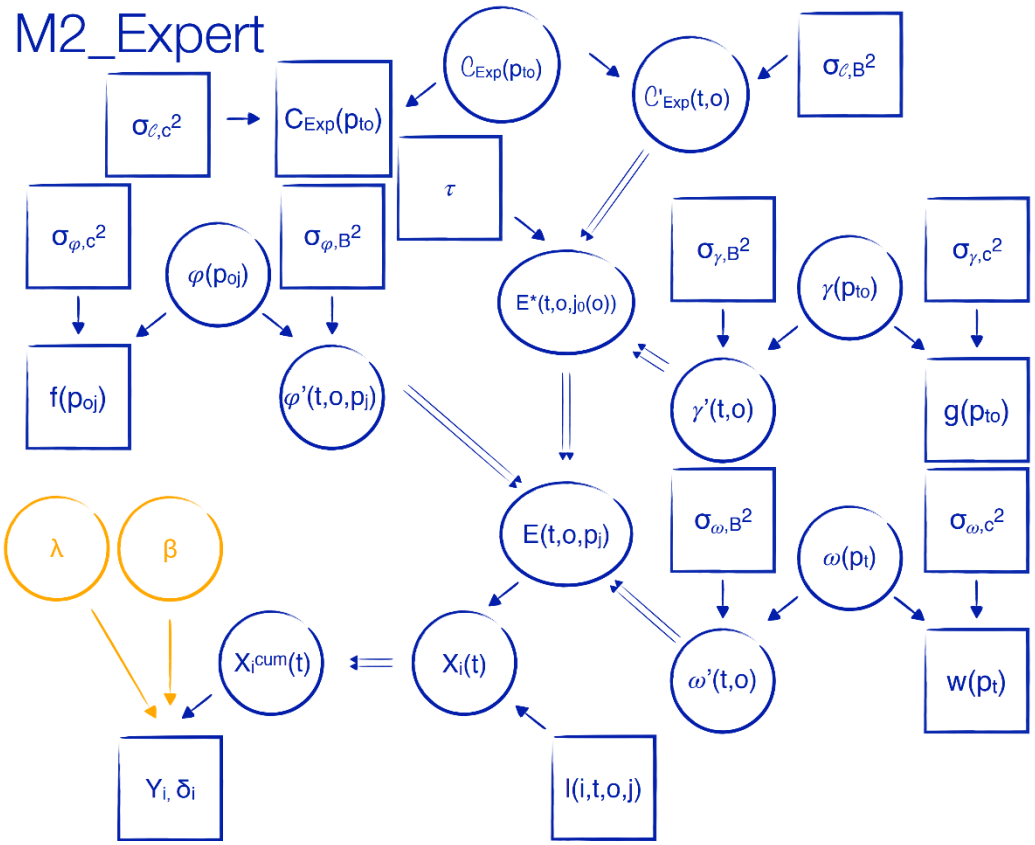

Figure 6.4: Hierarchical model combining a disease model with measurement model M2_Expert to describe how exposure uncertainty in underground mining objects in Saxony and Thuringia as well as in development objects in Saxony was accounted for on the data of the Wismut cohort for the second exposure assessment period in cases where radon gas concentration was determined by experts.

As already described, in cases where an object and/or year uses the exposures from a transfer object and/or transfer year, a transfer factor τ is multiplied for which no measurement error component is assumed. However, for measurement models M2, M2_Expert and M3, the respective transfer objects and transfer years are used to form clusters for which the classical error component of the mean annual radon concentration $U_{\mathcal{C},c}(p_{to})$ or $U_{\mathcal{C}_{Exp},c}(p_{to})$ is assumed to be shared. As described in *6.1.1*, M2, M3 and the new measurement model M2_Expert are modified so that a Berkson error component for the mean annual radon concentration $U_{\mathcal{C}',B}(t,o)$ or $U_{\mathcal{C}'_{Exp},B}(t,o)$ is now assumed for all years and objects belonging to a cluster, unless they are the reference year and object and do not themselves use transferred radon gas or radon progeny concentrations from another year or object. This modification would make less sense for M1a and M1b because the transfer of exposure values is somewhat different in these models. In particular, it involved the estimation of an actual transfer factor for which it makes sense to assume a classical error component. In contrast, in the transfer of exposure values in M2, M2_Expert and M3 the assumed transfer factor is implicitly set to 1 rather than being actually estimated. Because the exposure values were simply imputed based on radon gas or radon progeny concentration measurements from another year and object. In this way, it is also possible to model complicated dependence structures, in which, for example, object A uses the radon gas concentration values from transfer object B, but B itself takes the values from object C, where object C itself has no transfer object or year and is thus considered the reference object for this cluster.

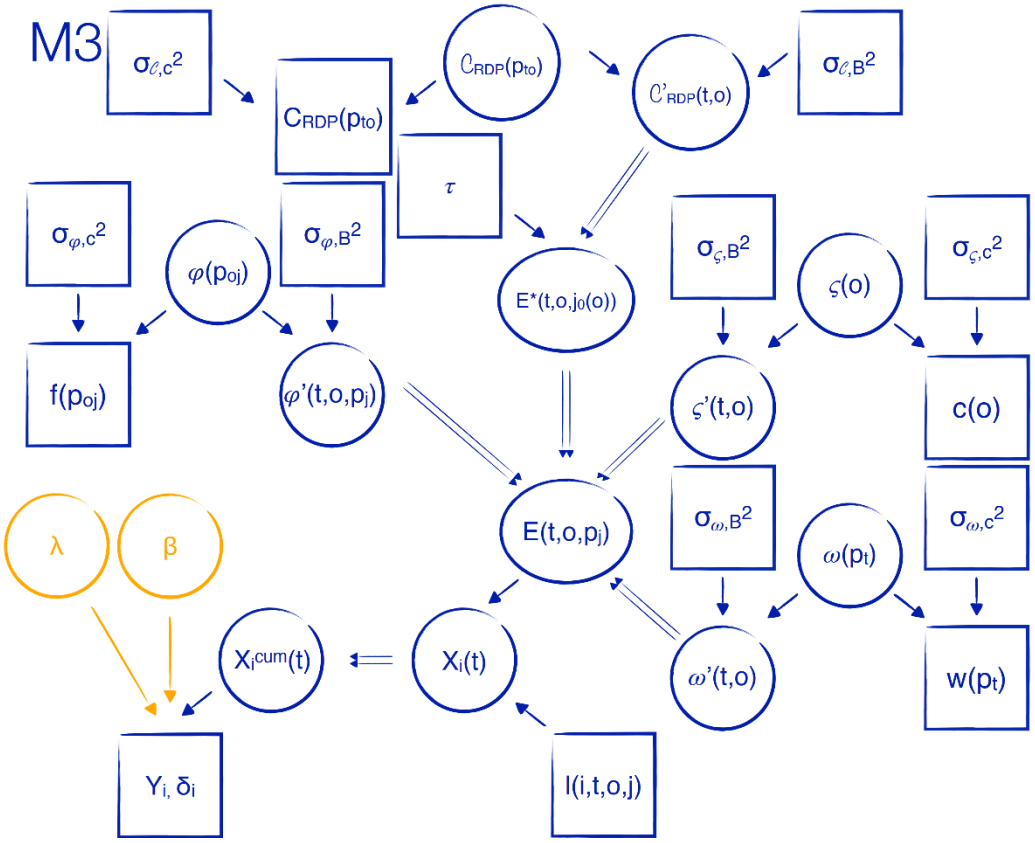

Figure 6.5: Hierarchical model combining a disease model with measurement model M3 to describe how exposure uncertainty in underground mining objects was accounted for on the data of the Wismut cohort for the third exposure assessment period.

In this example, all three objects A, B and C would be in one cluster for the classical error component, and the two objects A and B would each get an unshared Berkson error component.

Figure *6.3* and *6.5* show the adapted DAGs for underground mining objects for the second and third exposure assessment period in the Wismut cohort, respectively. Figure *6.4* shows the DAG for objects and years in the second exposure assessment period in which exposure assessment was not based on averaged measured radon gas concentration values, but the radon gas concentration was determined by experts.

M4

According to Chapter *2*, the measurement model M4 describes the uncertainties in the exposure assessment for surface areas affiliated to underground mining objects and exploration objects in Thuringia, whereby we assume a classical expert estimation error $U_{\mathcal{E},c}(p_{to})$ for the radon exposure determined by experts in WLM and for the activity weighting factor like always both a classical and a Berkson error component. Now we extend the scope of this measurement model to cases in which we lacked information to reconstruct the detailed exposure values according to M1a, M1b, M2, M3 or M6 or in which the exposure values in the extended cohort data could not be reproduced with these models. The measurement error model MX_Expert_WLM (M1a_Expert_WLM, M2_Expert_WLM, M3_Expert_WLM or M6_Expert_WLM, depending on the original measurement error model) was used if the radon exposure assessment was not clearly described in Lehmann et al. (1998) or Lehmann (2004), if the exposure assessment could not be reproduced with the defined models, or if the radon exposure assessment was too complicated or even impossible. The latter applied, for example, in cases where no information on the uranium recovering rate $r(t, o)$ was available for objects and years under M1a, which consequently makes a determination according to M1a impossible. Another example were cases where the associated mining object of an exploration

or development object was not clear, or consisted of a combination of different mining objects. In this case, it was not feasible to account for these complex dependence structures, which would result in some years and objects sharing a proportion of the classical error components of several years and objects at the same time. Finally, when the description by Lehmann et al. (1998) or Lehmann (2004) was very vague or did not agree at all with the exposure data, we preferred to choose model MX_Expert_WLM.

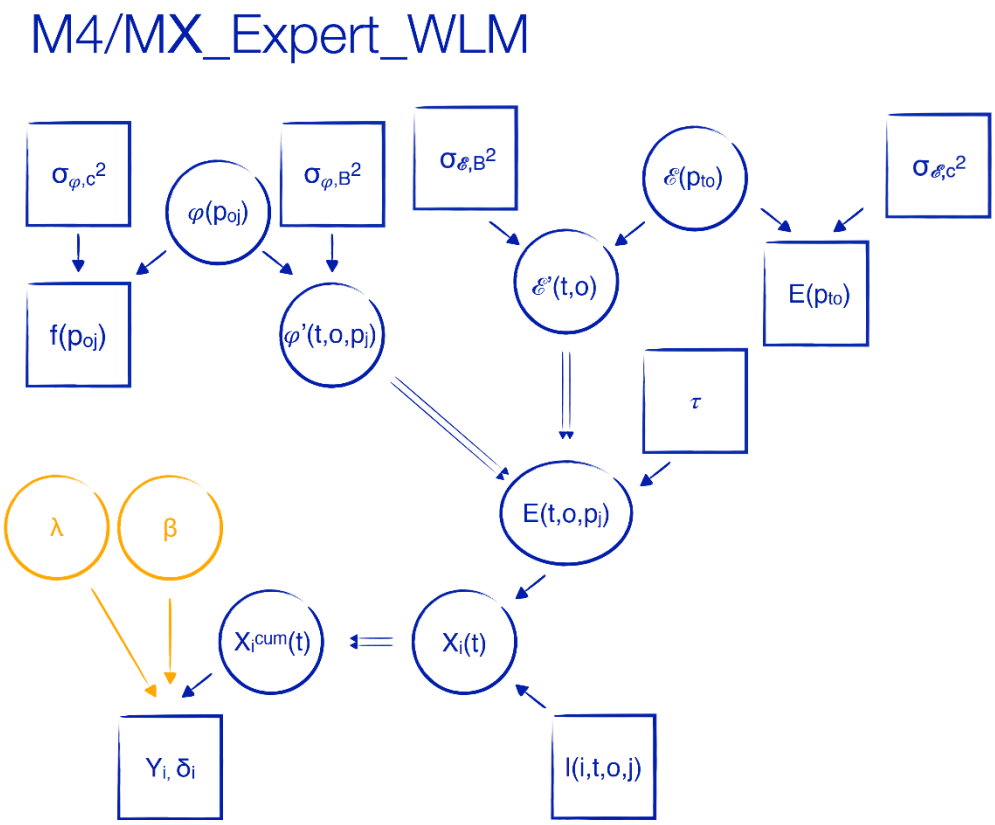

Figure 6.6: Hierarchical model combining a disease model with measurement model to describe how exposure uncertainty in surface areas affiliated to underground mining objects and exploration objects in Thuringia (M4) was accounted for on the data of the Wismut cohort. This measurement error model was also used in cases where radon exposure assessment was not clearly described, if exposure assessment could not be reproduced or if the radon exposure assessment was too complicated or impossible (MX_Expert_WLM).

As described for measurement models M2 and M3, it is also possible for measurement models M4 and MX_Expert_WLM to model complicated dependence structures via the transfer objects and transfer years resulting from the error-free transfer factor τ. For this purpose, clusters are again formed via the transfer objects and transfer years, for which the expert estimate of the exposure is then assumed to consist of a shared classical error component $U_{\mathcal{E},c}(p_{to})$ for all objects and years in the cluster, as well as an unshared Berkson error component $U_{\mathcal{E},B}(t,o)$ for all objects and years in the cluster, that are not the reference object or year of this cluster. Additionally, for M4 and MX_Expert_WLM, the years within an object which have the same annual radon progeny exposure value, are considered as one cluster, even if there is no transfer year specified in the data set. For such cases, a correspondingly shared classical error component $U_{\mathcal{E},c}(p_{to})$ and an additional unshared Berkson error component $U_{\mathcal{E},B}(t,o)$ are assumed for the expert estimation of the radon progeny exposure for each year.

Figure *6.6* shows the updated DAG for surface areas affiliated to underground mining objects and exploration objects in Thuringia and for other objects with expert estimation of the annual radon progeny exposure.

M5

Due to feasibility issues and in accordance with BfS, it was not within the scope of the current work to account for measurement error for exposures in processing companies and they were therefore excluded from the analysis. For this, all miners who had worked in a processing company even once in their career were completely excluded. Thus, the two measurement models for processing workers M5a and M5b are not relevant when accounting for exposure uncertainty on the extended data of the Wismut cohort for this project.

M6

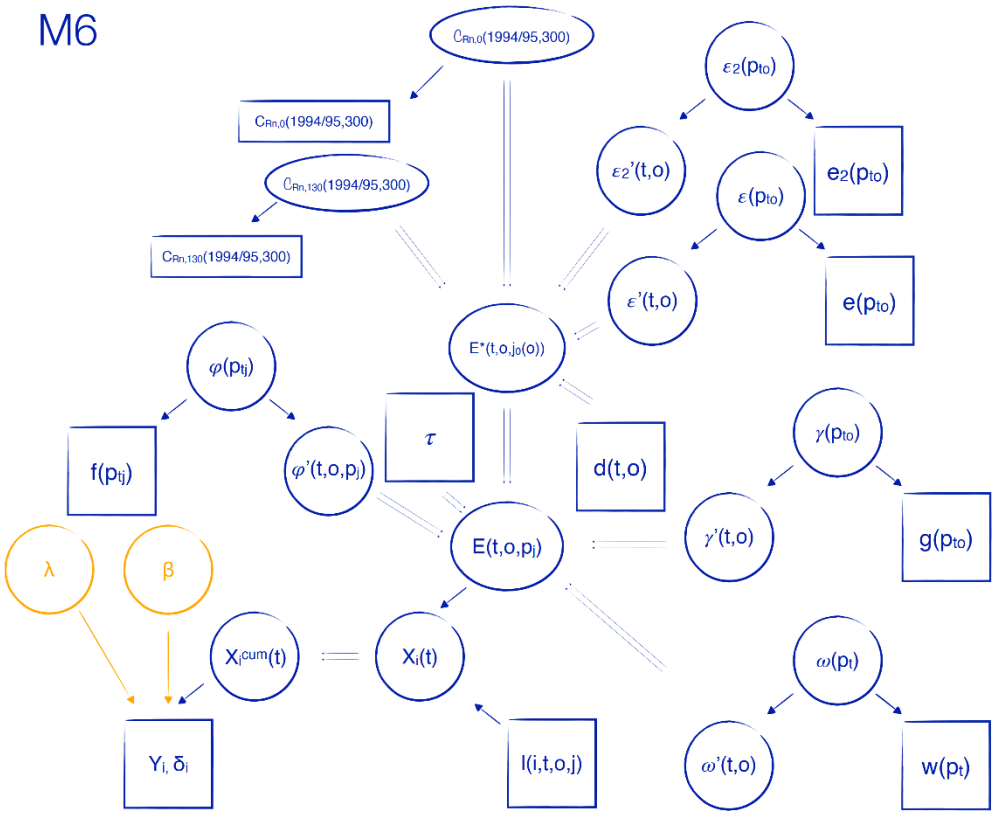

Figure 6.7: Hierarchical model combining a disease model with measurement model M6 to describe how exposure uncertainty for open pit mining objects was accounted for on the data of the Wismut cohort. Due to the limited space and for a clearer presentation, no measurement error variances are shown here.

Figure 6.7 shows the exposure assessment for open pit mining objects. In order to determine the annual radon exposure of an open pit mining object, in addition to the three parameters $g(p_{to})$, $w(p_t)$ and $f(p_{tj})$, the radon gas concentration measurements from 1994/1995 in object 300 Lichtenberg at a depth of 0 and 130 meters, and the two evaluation factors $e(p_{to})$ and $e_2(p_{to})$ as well as the depth $d(t,o)$ are necessary. However, the source in which, according to Lehmann et al. (1998), the depth for some years and open pit mining objects should be contained, was not available for this project. Thus, the depth $d(t,o)$ was calculated from the observed exposure values from the extended cohort data $Z_i(t,o)$ in a similar way to the evaluation area $A(t,o)$ from M1a as

$$d(t,o) = \left(\frac{Z_i(t,o)}{g(p_{to}) \cdot w(p_t) \cdot f(p_{tj})} \cdot \frac{3700}{12} - C_{Rn,0}(1994/1995{,}300)\right) \cdot \frac{130}{(C_{Rn,130}(1994/1995{,}300) - C_{Rn,0}(1994/1995{,}300)) \cdot e(p_{to}) \cdot e_2(p_{to})}$$

and as already described before in Chapter *2*, the depth was assumed to be known without measurement error (see Table *2.1*). Lehmann et al. (1998) gives the values for the evaluation factors

$e(p_{to})$ and $e_2(p_{to})$, whereby we assume that for years without explicit specification of a value, the value from the last specified year was used.
All objects and years with exposure assessment according to measurement model M6 are assumed to belong to one cluster with shared classical error for the mean radon gas concentrations of basic exposure without mining activity for object 300 Lichtenberg in 1994/1995 at ground level and in a depth of 130m.

In comparison to the DAG in Chapter *2*, the parameter τ is shown in Figure *6.7*.

6.1.4 Adapting the quantified values of measurement error variances

Since the adjusted measurement models sometimes have different uncertain parameters and associated error components compared to those from Chapter *2*, the quantified values of measurement error variances must also be adjusted. For all uncertain parameters from the adapted measurement models Table *6.1* lists their respective measurement models, their assumed error types (additive or multiplicative and Berkson or classical) and the assumed values for the standard deviations of the errors and an explanation of how these values were obtained. For the Berkson error component for $\mathcal{E}(p_{to})$, $\mathcal{C}_{Exp}(p_{to})$, $\mathcal{C}_{Rn}(p_{to})$ and $\mathcal{C}_{RDP}(p_{to})$, which results from the clustering due to transfer objects and years, the quantified value of the standard deviation of the Berkson error component of the transfer factor is used. We also adapted the Berkson error component for the working time factor $\omega(p_t)$, since the initially assumed standard deviation of 0.69 seemed disproportional compared with the standard deviation parameter for the classical error component of 0.04. We therefore adapted the Berkson error standard deviation to reflect the same proportion of the classical and Berkson standard deviations as for $\gamma(p_{to})$. Since the ratio of these parameters is $0.33, (0.23/0.69)$, we assumed a Berkson error standard deviation of 0.12.

6.1.5 Shaft specific exposure estimation and changes in object association

The assessment of the radon progeny exposure by the JEM 1 (Lehmann et al., 1998) is solely based on the different objects of the Wismut, whereas the JEM 2 (Lehmann, 2004) differentiates the radon exposure for some objects according to shafts or groups of shafts with sometimes significantly different exposure values. We always differentiate between the different shaft groups if the shaft specific exposure tables in Lehmann (2004) show a different exposure value than the one from the underlying object in Lehmann et al. (1998) for at least one year. This means that sometimes there are different shaft groups within an object, which have different radon exposures for some years but are based on the same exposure assessment for other years. In years where the radon exposure assessment is different, the shaft groups belong to different clusters for the shared error components, whereas they belong to the same cluster for years with identical exposure assessment. Although there were no shaft specific exposure documentations, but only the object specific exposures from Lehmann et al. (1998) for objects with an exposure estimation according to M4, we observed discrepancies between the reported values for radon progeny exposure in Lehmann et al. (1998) and the annual radon progeny exposure that were documented in the extended Wismut cohort data for some shafts. In these cases, we decided to assume shaft specific exposure values by estimating the corresponding radon progeny exposure value from the extended Wismut cohort data. Additionally, many objects and especially shafts have been renamed over time, or shafts have been assigned to other objects. As a consequence, it often happens that miners from the same shaft are listed under different shaft and object names. To ensure that the errors for these miners are shared, we renamed these shafts so that they all have the same shaft and object name, whereby the object corresponds to the respective object the shaft was assigned to for the given year. Accordingly, object names were also changed if an object was connected to another object, or if an object had already been closed and, according to Lehmann (2004), it could only be assumed that the miners were further employed in another object.

Table 6.1: Uncertain parameters with the measurement models in which they occur, their assumed error types and the quantified standard deviations as used when accounting for measurement error on the Wismut data as well as the quantification approach. Due to lack of space, we abbreviated 'classical error standard deviation' to 'classical standard deviation' and 'Berkson error standard deviation' to 'Berkson standard deviation'.

parameter	measurement model	assumed measurement error type	standard deviation	quantification approach
$\mathcal{E}(p_{to})$	M4, MX_Expert_WLM	$U_{\mathcal{E},c}$: classical multiplicative	0.936	Allodji et al. (2012b) for French cohort
		$U_{\mathcal{E}',B}$: Berkson multiplicative	0.18	quantified for Berkson error for the transfer factor in M4
$C_{Exp}(p_{to})$	M2_Expert	$U_{C_{Exp},c}$: classical multiplicative	0.936	Allodji et al. (2012b) for French cohort
		$U_{C'_{Exp},B}$: Berkson multiplicative	0.33	quantified for Berkson error for the transfer factor in M2
$C_{Rn}(p_{to})$	M2	$U_{C,c}$: classical additive	0.59	quantified for 009 Aue 1961
		$U_{C',B}$: Berkson multiplicative	0.33	quantified for Berkson error for the transfer factor in M2
$C_{RDP}(p_{to})$	M3	$U_{C,c}$: classical additive	0.03	quantified for 009 Aue 1968
		$U_{C',B}$: Berkson multiplicative	0.13	quantified for Berkson error for the transfer factor in M3
$C_{Rn}(t_0(o_0(o)), o_0(o))$	M1a, M1b	$U_{C,c}$: classical additive	5.29	quantified for 009 Aue 1955
$C_{Rn}(1937/38, 003)$	M1a	$U_{C,c}$: classical additive	6.56	quantified for 003 Schneeberg
$C_{Rn,0}(1994/95, 300)$, $C_{Rn,130}(1994/95, 300)$	M6	$U_{C,c}$: classical additive	0.59	quantified for 009 Aue 1961
$\omega(p_t)$	M1a, M1b, M2, M2_Expert, M3, M6	$U_{\omega,c}$: classical multiplicative	0.04	Allodji et al. (2012a) for French cohort
		$U_{\omega',B}$: Berkson multiplicative	0.12	equal proportions of the classical and Berkson standard deviations for $\omega(p_t)$ and $\gamma(p_{to})$
$\varphi(p_{tj})$	M1a, M1b, M2, M2_Expert, M3, M4, MX_Expert_WLM, M6	$U_{\varphi,c}$: classical multiplicative	0.33	classical standard deviation as quantified for the evaluation factor
		$U_{\varphi',B}$: Berkson multiplicative	0.69	Berkson standard deviation as quantified for $\gamma(p_{to})$
$\gamma(p_{to})$	M1a, M1b, M2, M2_Expert, M6	$U_{\gamma,c}$: classical multiplicative	0.23	quantified for classical error for $\gamma(p_{to})$
		$U_{\gamma',B}$: Berkson multiplicative	0.69	quantified for Berkson error for $\gamma(p_{to})$
$\varsigma(o)$	M3	$U_{\varsigma,c}$: classical multiplicative	0.33	classical standard deviation as quantified for the evaluation factor
		$U_{\varsigma',B}$: Berkson multiplicative	1.45	quantified for Berkson error for $\varsigma(o)$
$\mathcal{B}(o)$	M1a	$U_{\mathcal{B},c}$: classical multiplicative	0.33	classical standard deviation as quantified for the evaluation factor
		$U_{\mathcal{B}',B}$: Berkson multiplicative	0.69	Berkson standard deviation as quantified for $\gamma(p_{to})$
$\epsilon(p_{to})$, $\epsilon_2(p_{to})$	M6	$U_{\epsilon,c}$: classical multiplicative	0.33	quantified for classical error for the evaluation factor
		$U_{\epsilon',B}$: Berkson multiplicative	0.69	Berkson standard deviation as quantified for $\gamma(p_{to})$
$\tau_e(o)$	M1a, M1b	$U_{\tau_e,c}$: classical multiplicative	0.37	quantified for classical error for the transfer factor
		$U_{\tau'_e,B}$: Berkson multiplicative	0.33	quantified for Berkson error for the transfer factor in M2

6.1.6 Calculating exposure values for individual miners

The estimated exposures according to the measurement models or the ones given in the JEMs (Lehmann et al., 1998; Lehmann, 2004) correspond to the annual radon progeny exposure, however, it frequently occurred that a miner worked less than a whole year in an object, e.g. because he started or ended his working career at the Wismut in the middle of the year, or because he worked in different objects or in different activities within a year. In these cases, it would be wrong to consider the whole annual radon exposure for this year t, so we weight the annual radon exposure by the percentage of days $l(i,t,o,j)$ that miner i worked at that object o or with that activity j. In the extended cohort data set the number of days is given in increments of 10, which in combination with different objects during a year can result in a maximum number of 380 days per year for individual miners. Thus, we calculate the individual working history of miner i as $l(i,t,o,j) =$ min(proportion of days in object o with activity j, 1).

As described in Section *4.4*, we need to account for time-varying exposure in proportional hazards models to avoid the so-called time-dependent bias (van Walraven et al., 2004; Beyersmann et al., 2008; Wolkewitz et al., 2012; Barnett et al., 2011). In its original form, the data are available in a so-called wide format in which one line represents a miner and the information on yearly exposure values is given through different columns. To account for time-varying exposure, in the extended Wismut cohort data, the follow-up time of the miners is partitioned into intervals at which the cumulative exposure was constant. This is achieved by adding a row for each time interval at which the exposure of a miner remained constant, resulting in a long format. For the additional rows in the data set that are thereby created (Therneau and Grambsch, 2000; Therneau and Crowson, 2013) and for all other rows, we have to specify at which exact point in time the new cumulative exposure value started and when it stopped. As we only know the annual exposure values to radon progeny, we have to choose a date on which the miner received this annual exposure every year. We chose the 1st of January of each year, implying a step function with jumps on the 1st of January in every year that a miner received an exposure. While it might have been more reasonable to assume that a miner received the annual exposure value on the 30th of June or on the 31st of December to not overestimate his exposure value at any given time, it was more convenient to choose the 1st of January as some miners worked part of the year in one object and part of the year in another. In these situations, there is not only one additional row in the data set for a given year of a miner, but we create several rows in the data set for a given miner and year that define the time intervals at which the miner worked in these different objects to be able to weight the corresponding values by the proportion of days the miner worked in each object (as described above). Given these rows for miners working in several objects in a given year, it is more logical to distribute the number of days in the working history of a miner between the 1st of January and the 31st of December rather than distributing it between the 30th of June and the 30th of June of the following year or between the 31st of December and the 31st of December of the following year. In accordance with common practice in the analysis of lung cancer mortality in cohorts of uranium miners, we exclude exposure values that a miner received immediately before death by lung cancer by lagging cumulative exposure $X_i^{\text{cum}}(t)$ to respect a latency period of five years (Grosche et al., 2006; Amabile et al., 2009; Hauptmann et al., 2001; Langholz et al., 1999; Richardson et al., 2011).

6.2 Applying the Bayesian hierarchical approach to the Wismut cohort data

The results were generated using the implemented MCMC algorithm. We run eight chains with 100000 iterations and 50000 iterations as burn-in. We only keep each 200th iteration (thinning) to reduce autocorrelation in each chain. Before starting the burn-in phase, we tune the acceptance rates by using 100 adaptive phases. For that, we run a small number of iterations 100 times and look at the acceptance rates. If the rates are not sufficient, we adapt the standard deviations of the proposal distributions to get better acceptance rates.
For measurement models M2, M2_Expert and M3, we did not observe enough flexibility when

defining only one prior distribution for $\mathcal{C}_{Rn}, \mathcal{C}_{Exp}$ and $\mathcal{C}_{RPD}$ as the variation was too high between the years. We therefore defined one individual distribution for each year. This means, for instance in measurement model M2, that instead of one $\mu_{\mathcal{C}_{Rn}}$ and $\sigma_{\mathcal{C}_{Rn}}$, we actually have two parameters for every year from 1955 to 1974. However, we will not use a specific notation for that since each of them is specified with the same prior.

Prior parameters

Here, we present the used prior parameters in the algorithm. For the parameters of interest, we specify priors as follows:

- $\beta \sim N(0{,}100)$ (normal distribution parameterized with mean and standard deviation)
- $\lambda_1 \sim Ga(600{,}1/10000000)$ (Gamma distribution parameterized with shape and scale)
- $\lambda_2 \sim Ga(12000{,}1/1000000)$
- $\lambda_3 \sim Ga(46000{,}1/1000000)$
- $\lambda_4 \sim Ga(1000{,}1/100000)$

Parameters of the exposure models are specified as

- $\mathcal{C}_{Rn_{old}} \sim N^+(22.5{,}4)$ (normal distribution with values only on $\mathbb{R}^+$; only for M1a)
- $\mathcal{C}_{Rn_{ref}} \sim N^+(34.09{,}10)$ (only for M1a)
- $b \sim B(1{,}1)$ (beta distribution with parameters $a = b = 1$; only for M1a), truncated at $[0.15{,}1.1]$
- $\tau_e \sim B(1{,}1)$ (only for M1a) truncated at $[0.3{,}1]$
- $\mathcal{C}_{Rn} \sim N^+(\mu_{\mathcal{C}_{Rn}}, \sigma_{\mathcal{C}_{Rn}})$ (only for M2)
- $\mathcal{C}_{Exp} \sim N^+(\mu_{\mathcal{C}_{Exp}}, \sigma_{\mathcal{C}_{Exp}})$ (only for M2_Expert)
- $\mathcal{C}_{RPD} \sim N^+(\mu_{\mathcal{C}_{RPD}}, \sigma_{\mathcal{C}_{RPD}})$ (only for M3)
- $\zeta \sim B(1{,}1)$ (only for M3) truncated at $[0.15{,}1.7]$
- $\varphi \sim B(1{,}1)$ truncated at $(0{,}1.3]$
- $\omega \sim B(1{,}1)$ truncated at $[0.6{,}1.5]$
- $\gamma \sim B(1{,}1)$ (only for M1a, M2, M2_Expert) truncated at $[0.05{,}0.8]$

For C_{Rn}, C_{RPD} and C_{Exp} we specify hierarchical priors:

- $\mu_{\mathcal{C}_{Rn}} \sim N(6{,}10)$ (one distribution for each year in M2)
- $\sigma_{\mathcal{C}_{Rn}} \sim N(8{,}10)$ (one distribution for each year in M2)
- $\mu_{\mathcal{C}_{Exp}} \sim N(0.15{,}0.2)$ (one distribution for each year in M2_Expert)
- $\sigma_{\mathcal{C}_{Exp}} \sim N(0.2{,}0.2)$ (one distribution for each year in M2_Expert)
- $\mu_{\mathcal{C}_{RPD}} \sim N(1.7{,}3)$ (one distribution for each year in M3)
- $\sigma_{\mathcal{C}_{RPD}} \sim N(1{,}2)$ (one distribution for each year in M3)

The prior specification is the same for the 1960+ and the full cohort, except that for the 1960+ cohort we omit all priors that occur only for measurement model M1a.

6.3 Results

In the following, we present the results for the 1960+ cohort and the full cohort when accounting for measurement error. These results are preliminary, as they only account for a subset of the 1960+ cohort data and their robustness still has to be confirmed in extensive sensitivity analyses. Moreover, they do not include data on workers who were employed in Wismut processing companies at any point during their working career.

6.3.1 Results for the 1960+ cohort

Table 6.1 shows the results with measurement error correction on the sub-cohort of miners hired in 1960 or later (also referred to as 1960+ cohort (Kreuzer et al., 2018)) when accounting for measurement error models M2, M3 and M2_Expert as described in Section *6.1*. Table *6.2* shows the results without measurement error correction.

Table 6.1: Results when correcting for measurement models M2, M3 and M2_Expert on the sub-cohort of miners hired in 1960 or later.

parameter	**Estimate [95% Credible Interval]**
EHR (beta)	1.80 [0.70; 3.36] per 100 WLM
Baseline hazard ($0 < \text{age} \leq 40$)	2.44 [1.09; 4.03] in 10^{-5}
Baseline hazard ($40 < \text{age} \leq 55$)	2.44 [1.87; 3.16] in 10^{-4}
Baseline hazard ($55 < \text{age} \leq 75$)	1.81 [1.52; 2.12] in 10^{-3}
Baseline hazard ($75 < \text{age} \leq 104$)	5.94 [3.45; 8.97] in 10^{-3}

Table 6.2: Results when not correcting for measurement error the sub-cohort of miners hired in 1960 or later.

Parameter	**Estimate [95% Credible Interval]**
EHR (beta)	1.44 [0.73; 2.52] per 100 WLM
Baseline hazard ($0 < \text{age} \leq 40$)	2.43 [1.19; 4.07] in 10^{-5}
Baseline hazard ($40 < \text{age} \leq 55$)	2.40 [1.79; 3.08] in 10^{-4}
Baseline hazard ($55 < \text{age} \leq 75$)	1.79 [1.48; 2.12] in 10^{-3}
Baseline hazard ($75 < \text{age} \leq 104$)	5.99 [3.47; 9.50] in 10^{-3}

We included miners who had only been exposed in years and objects that were characterized through these three measurement models, i.e. we excluded all miners who had been exposed in any year and object characterized through measurement models M1a, M1b, M4, MX_Expert_WLM, M5 or M6 in the 1960+ cohort resulting in a reduction to 18852 workers and 286 deaths by lung cancer (compared to the 23899 workers and 458 total number of deaths by lung cancer observed in the 1960+ cohort). While the estimates of the baseline hazard parameters remain largely unchanged through the correction of measurement error, we estimate an EHR of 1.80 per 100 WLM, i.e. for each increase of 100 WLM the hazard rate is increased by 1.80. Thus, it is larger than the uncorrected EHR estimate of 1.44. This increase is accompanied by larger uncertainty intervals. Note also that the measurement error corrected 95% credible interval [0.70; 3.36] per 100 WLM encompasses the uncorrected interval [0.73; 2.52] with both intervals excluding the 0.

6.3.2 Results for the full cohort

Table *6.3* shows the results with measurement error correction on the full cohort when accounting for measurement error models M1a, M2, M3 and M2_Expert as described in Section *6.1*. Table *6.4* shows the results without measurement error correction on the full cohort.

Table 6.3: Results when correcting for a measurement error on models, M1a, M2, M3 and M2_Expert on the full cohort.

parameter	**Estimate [95% Credible Interval]**
EHR (beta)	0.54 [0.35; 0.81] per 100 WLM
Baseline hazard ($0 < \text{age} \leq 40$)	2.26 [1.25; 3.72] in 10^{-5}
Baseline hazard ($40 < \text{age} \leq 55$)	3.61 [3.07; 4.17] in 10^{-4}
Baseline hazard ($55 < \text{age} \leq 75$)	2.40 [2.18; 2.63] in 10^{-3}
Baseline hazard ($75 < \text{age} \leq 104$)	4.09 [3.51; 4.64] in 10^{-3}

Table 6.4: Results when not correcting for a measurement error on models, M1a, M2, M3 and M2_Expert on the full cohort.

parameter	**Estimate [95% Credible Interval]**
EHR (beta)	0.33 [0.27; 0.4] per 100 WLM
Baseline hazard ($0 < \text{age} \leq 40$)	2.33 [1.29; 3.75] in 10^{-5}
Baseline hazard ($40 < \text{age} \leq 55$)	3.67 [3.16; 4.24] in 10^{-4}
Baseline hazard ($55 < \text{age} \leq 75$)	2.44 [2.22; 2.65] in 10^{-3}
Baseline hazard ($75 < \text{age} \leq 104$)	4.06 [3.43; 4.75] in 10^{-3}

We included miners who had only been exposed in years and objects that were characterized through these four measurement models, i.e. we excluded all miners who had been exposed in any year and object characterized through measurement models M1b, M4, MX_Expert_WLM, M5 or M6 resulting in a reduction to 30271 workers and 1302 deaths by lung cancer (compared to the 50470 workers and 3438 total number of deaths by lung cancer observed in the full cohort). If we include data on all 50470 workers, we obtain an uncorrected EHR of 0.23 [0.21; 0.26] per 100 WLM. Again, the estimates for baseline hazard parameters remain largely unchanged whereas we can observe a noticeable increase in the EHR when correcting for measurement error that is accompanied by larger uncertainty intervals. We estimate an EHR of 0.54 per 100 WLM, i.e. for each increase of 100 WLM the hazard rate is increased by 0.54.

6.3.3 Stability of the results regarding the assumed magnitude of the assumed exposure uncertainty

In order to assess the stability of the estimated parameters as a function of the assumed magnitude of the exposure uncertainty, we ran a sensitivity analysis in which we additionally fitted four models in which we systematically varied the assumed magnitude of the measurement errors by multiplying

the magnitude of all Berkson error components or all classical error components by a factor of 0.5 or 1.5 in order to assess how increasing or reducing the magnitude of the measurement error components by 50% would affect the results. We preferred a 50% increase rather than doubling the magnitude of measurement error components, because we were concerned that a doubling of the magnitude might result in highly implausible values. This sensitivity analysis was only performed for the 1960+ cohort. As can be seen in Figure *6.8*, both the risk estimates and the width of the 95% credible intervals strongly depends on the assumed magnitude of Berkson and classical error components. In line with the results of the simulation study shown in Section 5.7, we see that assuming a larger magnitude of measurement error can lead to substantial increases in the EHR. Interestingly, we only observe this phenomenon when we increase the Berkson components of the different uncertain parameters, whereas increasing the classical error components by 50% leads to a reduction in the EHR estimate that is close to the uncorrected estimate. Conversely, assuming a smaller magnitude of classical error does not substantially change the EHR estimate and credible intervals whereas assuming a smaller magnitude of Berkson error leads as expected to a smaller EHR estimate. Overall, the results of this sensitivity analysis underline the importance of performing an extensive quantification of exposure uncertainty as the resulting EHR estimates are highly contingent upon the assumed magnitude of Berkson and classical measurement error.

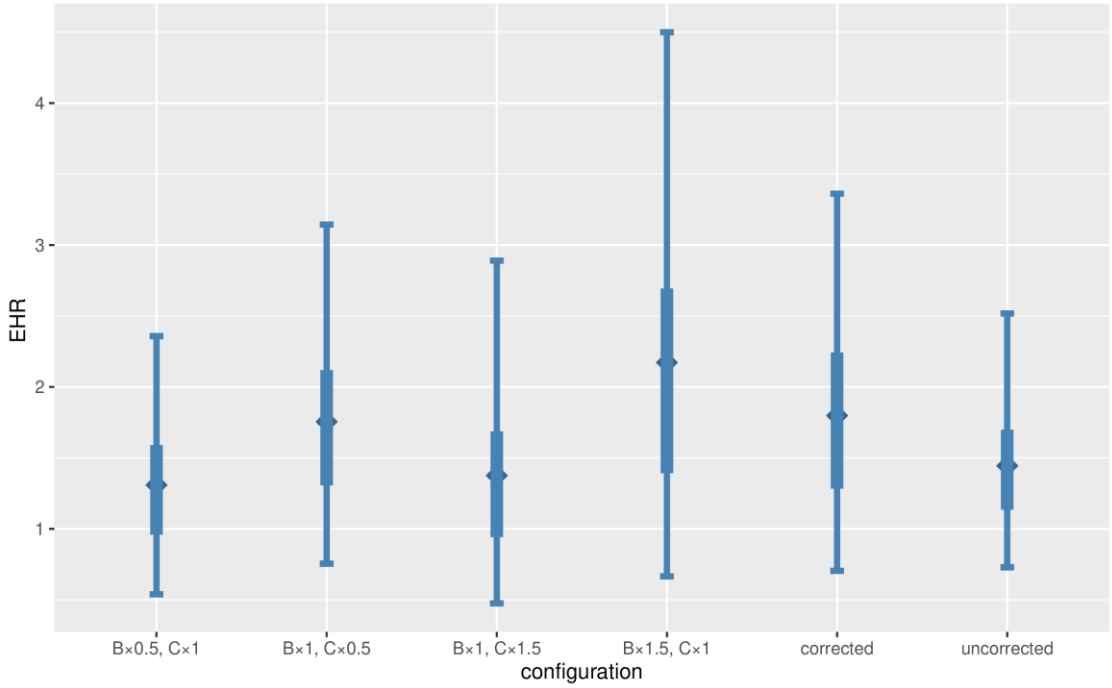

Figure 6.8: Estimate of the EHR with a 50% and 95% credible interval for the different settings. B denotes a Berkson error and C a classical error. For example, B×0.5, C×1 means that the standard errors for the Berkson errors are assumed to be only half of the value assumed for the main analysis and the standard error of the classical error is on its original value assumed for the main analysis. 'corrected' and 'uncorrected' refer to the results from the main analysis and serve as a reference.

7 Discussion

In this work, the aim was to quantify exposure uncertainty in the Wismut cohort, to derive measurement models to describe the exposure uncertainty in this cohort and to develop an approach to correct for measurement error. The performance of the proposed approach was assessed on simulated data and compared with classical approaches to account for measurement error. The proposed approach was applied to the data of the Wismut cohort (mortality follow-up 1946-2013) without accounting for effect modifying variables. Workers who were employed in processing companies at any point during their working career were excluded due to feasibility issues.

7.1 The applicability of the proposed approach

This work demonstrated that the developed Bayesian hierarchical approach is very suitable for the correction of exposure uncertainty in the Wismut cohort and that it can be applied to account for a wide range of measurement error structures. Due to its flexibility, it was possible to account for complex structures of exposure uncertainty with measurement errors in radon gas and radon progeny concentrations and parameter uncertainties in the estimates of various uncertain quantities including the working time factor, the activity weighting factor, the equilibrium factor and the ventilation correction factor. It was possible to account for changes in the structure and the magnitude of measurement error that occurred due to changes in the methods of exposure assessment over time and for dependence structures both for several exposure years and for several objects. Based on these results, we can see that it is very suitable and applicable for the correction of measurement error in the Wismut cohort, and more generally, to account for complex structures of measurement error in occupational cohorts.

7.2 Comparing the results with previous findings from the literature

In the current work, we estimated an EHR of 1.44 per 100 WLM with a 95% credible interval of [0.73; 2.52] on the sub-cohort of miners hired in 1960 or later without measurement error correction when modelling the association between time until death by lung cancer and radon exposure through a proportional hazards model. Kreuzer et al. (2018) used a grouped Poisson regression model on the data based on the same mortality follow-up (1946-2013) and estimated an Excess Relative Risk per 100 WLM of 1.1 [0.6; 1.7] on the sub-cohort of miners hired in 1960 or later. The difference in the uncorrected risk coefficients might be explained by the fact that we used a proportional hazards model instead of a grouped Poisson model and that we had to exclude all miners who were ever exposed in any year or object characterized through measurement models M4, MX_Expert_WLM, M5 or M6 on this cohort, resulting in a decrease of lung cancer deaths from 495 to 286 events. When accounting for measurement error, the EHR and 95% credible interval increased to 1.80 per 100 WLM with a 95% credible interval of [0.70; 3.36]. The increase in the risk estimate and the increased uncertainty around this estimate are in line with the results of the simulation study that we conducted to assess the performance of the proposed approach and with results on the French cohort of uranium miners where the EHR and 95% credible intervals increased from 0.88 [0.50; 1.36] to 0.90 [0.51; 1.41] when correcting for measurement error (Hoffmann, 2017). Note that the difference in naive risk estimates between the Wismut cohort and the French cohort of uranium miners is in line with previous comparisons of the risk estimates in these two cohorts (Tirmarche et al., 2012). The current results for the Wismut cohort are preliminary, as they do not yet account for all measurement models that are to be considered. Moreover, in future work, it seems essential to consider effect modifying variables in the association between radon exposure and lung cancer mortality in the Wismut cohort. On the one hand, previous simulation studies suggest that complex structures of measurement error may lead to apparent effect modification when the true model is linear without effect modification (Hoffmann, 2018b). On the other hand, if the true model includes

effect modifying variables, the correction of measurement error is unlikely to be reliable because the resulting misspecification in the disease model provides erroneous information in the updating of the latent variables.
Finally, these results rely on many assumptions concerning the structure, type and magnitude of measurement error that we will discuss in the following paragraph and their robustness still has to be confirmed in extensive sensitivity analyses.

7.3 Assumptions on the structure, type and magnitude of measurement error

Due to the absence of any type of validation, calibration or replication study that would allow us to obtain ancillary information to assess the structure and magnitude of measurement error in the Wismut cohort, we had to make a great number of assumptions. In particular, we assumed an additive measurement error structure for the generalization error despite a common consensus in the literature that measurement error in radiation exposure in general, and in radon exposure in particular, is best described by a multiplicative measurement error component following a lognormal distribution (Lubin et al., 1995b; Stram et al., 1999; Heid, 2002; Heid et al., 2002; Heid et al., 2004; Heidenreich et al., 2004; Lubin et al., 2005; Advisory Group on Ionising Radiation AGIR, 2009; Heidenreich et al., 2012; Allodji et al., 2012a,b,c). The results put the assumption of additive and normally distributed generalization error somewhat in question because some of the mean parameters in the exposure model for radon gas and radon progeny concentrations were estimated to take negative values. These results indicate that the model tries to describe the very heavy tailed distribution of mean radon gas and mean radon progeny measurements through a truncated normal distribution even though such a skewed distribution would probably be better described through a lognormal distribution. In future work, it would be worthwhile to conduct sensitivity analyses in which a multiplicative measurement error structure is assumed.
Due to a lack of information, it was not possible to quantify the classical and the Berkson component for all uncertain quantities individually in the present work. We therefore made the assumption that the standard deviation quantified for classical error component of the evaluation could be transferred from this uncertain parameter to other uncertain parameters. This assumption is somewhat questionable and it is advisable to assess the robustness of the results to this assumption in future sensitivity analyses. Finally, in the present work, the value quantified for object 009 Aue in 1961 was always used for the standard deviation for the classical error components for $\mathcal{C}_{Rn}(p_{to})$, $\mathcal{C}_{Rn,0}(1994/95{,}300)$ and $\mathcal{C}_{Rn,130}(1994/95{,}300)$. Similarly, the value quantified for object 009 Aue in 1968 was used for the standard deviation for the classical error component for $\mathcal{C}_{RDP}(p_{to})$ for all years and objects belonging to measurement model M3. In future work, it would be worthwhile to assume different measurement error variances for the different years and different objects in a given measurement model for the generalization error in radon gas concentration measurements C_{Rn} and radon progeny concentration measurements C_{RPD} by accounting for the number of measurements that were averaged to obtain the final estimates.

7.4 Outlook

While the Bayesian hierarchical approach shows great flexibility when it comes to accounting for complex patterns of measurement error, the results of the current work have to be interpreted with caution. Indeed, they rely on many assumptions concerning the magnitude and the structure of measurement error. It was not in the scope of this project to conduct extensive sensitivity analyses to assess the robustness of the results to assumptions on the type and magnitude of measurement error in the Wismut cohort. In this context, it would also be valuable to make extensive posterior predictive checks to assess the fit of the model to the observed data. Thereby, it would be possible to assess whether time until death by lung cancer is systematically over- or underestimated for certain years and objects.
Finally, it was not in the scope of the current work to correct for measurement models M1b, M4,

MX_Expert_WLM, M5 and M6 in the Wismut cohort and it would be valuable to correct for these additional measurement error structures in future work to be able to provide a corrected risk estimate for all workers included in the 1960+ and in the full cohort.

References

[1] Advisory Group on Ionising Radiation AGIR (2009). Radon and Public Health. Technical report, Health Protection Authority.

[2] Allodji, R. S., Leuraud, K., Bernhard, S., Henry, S., Bénichou, J., and Laurier, D. (2012a). Assessment of Uncertainty Associated with Measuring Exposure to Radon and Decay Products in the French Uranium Miners Cohort. *Journal of Radiological Protection*, 32(1):85–100.

[3] Allodji, R. S., Leuraud, K., Thiébaut, A. C., Henry, S., Laurier, D., and Bénichou, J. (2012b). 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(2):151–163.

[4] Allodji, R. S., Schwartz, B., Diallo, I., Agbovon, C., Laurier, D., and de Vathaire, F. (2015). Simulation–Extrapolation Method to Address Errors in Atomic Bomb Survivor Dosimetry on Solid Cancer and Leukaemia Mortality Risk Estimates, 1950–2003. *Radiation and environmental biophysics*, 54(3):273–283.

[5] Allodji, R. S., Thiébaut, A., Leuraud, K., Rage, E., Henry, S., Laurier, D., and Bénichou, J. (2012c). 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*, 31(30):4428–4443.

[6] Althubaiti, A. and Donev, A. (2016). Non-Gaussian Berkson Errors in Bioassay Data. *Statistical Methods in Medical Research* 25 (1): 430–445.

[7] Amabile, J.-C., Leuraud, K., Vacquier, B., Caër-Lorho, S., Acker, A., and Laurier, D. (2009). Multifactorial Study of the Risk of Lung Cancer Among French Uranium Miners: Radon, Smoking and Silicosis. *Health Physics,* 97 (6): 613–621.

[8] Andrieu, C., Doucet, A., and Robert, C. P. (2004). Computational Advances for and from Bayesian Analysis. *Statistical Science,* 19 (1): 118–127.

[9] Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 72: 269–342.

[10] Augustin, T., Döring, A., and Rummel, D. (2008). 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*, 253–278. Springer.

[11] Austin, P. C. (2012). Generating Survival Times to Simulate Cox Proportional Hazards Models with Time-Varying Covariates. *Statistics in Medicine,* 31: 3946–3958.

[12] Barnett, A. G., Beyersmann, J., Allignol, A., Rosenthal, V. D., Graves, N., and Wolkewitz, M. (2011) The Time-Dependent Bias and Its Effect on Extra Length of Stay Due to Nosocomial Infection. *Value Health,* 14 (2): 381–386.

[13] Bartell, S. M., Hamra, G. B., and Steenland, K. (2017). Bayesian Analysis of Silica Exposure and Lung Cancer Using Human and Animal Studies. *Epidemiology,* 28 (2): 281–287.

[14]Bartlett, J. W., and Keogh, R. H. (2016). Bayesian Correction for Covariate Measurement Error: A Frequentist Evaluation and Comparison with Regression Calibration. *Statistical Methods in Medical Research*.

[15]Bender, R., Augustin, T., and Blettner, M. (2005). Generating Survival Times to Simulate Cox Proportional Hazards Models. *Statistics in Medicine,* 24 (11): 1713–1723.

[16]Besag, J., and Green, P. J. (1993). Spatial Statistics and Bayesian Computation. *Journal of the Royal Statistical Society: Series B (Methodological),* 55 (1): 25–37.

[17]Beyersmann, J., Gastmeier, P., Wolkewitz, M., and Schumacher, M. (2008). An Easy Mathematical Proof Showed That Time-Dependent Bias Inevitably Leads to Biased Effect Estimation. *Journal of Clinical Epidemiology,* 61 (12): 1216–1221.

[18]Blair, A., Thomas, K., Coble, J., Sandler, D. P., Hines, C. J., Lynch, C. F., Knott, C., Purdue, M. P., Zahm, S. H., Alavanja, M. C. R., Dosemeci, M., Kamel, F., Hoppin, J. A., Freeman, L. B., and Lubin, J. H. (2011). Impact of Pesticide Exposure Misclassification on Estimates of Relative Risks in the Agricultural Health Study. *Occupational and Environmental Medicine,* 68 (7): 537–541.

[19]Box, G. E. (1976). Science and Statistics. *Journal of the American Statistical Association,* 71 (356): 791–799.

[20]Brandl, G. (2021). Sphinx Documentation. *URL Http://Sphinx-Doc. Org/Sphinx. Pdf*.

[21]Brooks, S. P. (2003). Bayesian Computation: A Statistical Revolution. *Philosophical Transactions of the Royal Society of London,* 361: 2681–2697.

[22]Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC press.

[23]Buonaccorsi, J. P. (2010). Measurement Error - Models, Methods and Applications. Chapman Hall/CRC.

[24]Buonaccorsi, J. P., Dalen, I., Laake, P., Hjartaker, A., Engeset, D., and Thoresen, M. (2015). Sensitivity of Regression Calibration to Non-Perfect Validation Data with Application to the Norwegian Women and Cancer Study. *Statistics in Medicine,* 34.

[25]Buzas, J. S. (1998). Unbiased Scores in Proportional Hazards Regression with Covariate Measurement Error. *Journal of Statistical Planning and Inference*, 247–257.

[26]Buzas, J. S., Stefanski, L. A., and Tosteson, T. D. (2014). Measurement Error. In *Handbook of Epidemiology*, 1241–1282.

[27]Carroll, R. J. (2005). Measurement Error in Epidemiologic Studies. In *Encyclopedia of Biostatistics*. Vol. 5. Wiley Online Library.

[28]Carroll, R. J., Küchenhoff, H., Lombard, F., and Stefanski, L. A. (1996). Asymptotics for the SIMEX Estimator in Nonlinear Measurement Error Models. *Journal of the American Statistical Association,* 91 (433).

[29]Carroll, R. J., Ruppert, D., Crainiceanu, C. M., Tosteson, T. D., and Karagas, M. R. (2004). Nonlinear and Nonparametric Regression and Instrumental Variables. *Journal of the American Statistical Association,* 99 (467): 736–750.

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

[31]Carroll, R. J., and Stefanski, L. A. (1997). Asymptotic Theory for the SIMEX Estimator in Measurement Error Models. In *Advances in Statistical Decision Theory and Applications*, 151–164. Springer.

[32]Congdon, P. (2006). *Bayesian Statistical Modelling*. John Wiley & Sons.

[33]Cook, J. R., and Stefanski, L. A. (1994). Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association,* 89 (428): 1314–1328.

[34]Cox, D. R. (1975). Partial Likelihood. *Biometrika* 62 (2): 269–276.

[35]Cox, D. R. (1972). Regression Models and Life-Tables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 34 (2): 187–220.

[36]Eigenwillig, G. (2011). Der Uranerzbergbau Im Erzgebirge - Die Dadurch Bedingten Strahlenexpositionen Und Erkrankungen Der Bergleute: Eine Kritische Bewertung. Selbstverlag.

[37]Eigenwillig, G., and Ettenhuber, E. (2000). Strahlenexposition Und Strahleninduzierte Berufskrankheitenim Uranbergbau Am Beispiel Wismut. 3. Und Erweiterte Ausgabe. TÜV-Verl.:Köln.

[38]Ferrari, P., Carroll, R. J., Gustafson, P., and Riboli, E. (2008). 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,* 27: 6037–6054.

[39]Fischer, H. J., Vergara, X. P., Yost, M., Silva, M., Lombardi, D. A., and Kheifets, L. (2017). Developing a Job-Exposure Matrix with Exposure Uncertainty from Expert Elicitation and Data Modeling. *Journal of Exposure Science and Environmental Epidemiology,* 27 (1): 7–15.

[40]Flegal, K. M., Keyl, P. M., and Nieto, F. J. (1991). Differential Misclassification Arising from Nondifferential Errors in Exposure Measurement. *American Journal of Epidemiology,* 134 (10): 1233–1246.

[41]Gelman, A., and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 457–472.

[42]Geman, S., and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 721–741.

[43]Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Chapman Hall, Boca Raton.

[44]Green, P. J., Łatuszyński, K., Pereyra, M., and Robert, C. P. (2015). Bayesian Computation: A Summary of the Current State, and Samples Backwards and Forwards. *Statistics and Computing,* 25 (4): 835–862.

[45]Greenland, S. (2000). Principles of Multilevel Modelling. *International Journal of Epidemiology,* 29 (158–167).

[46]Grosche, B., Kreuzer, M., Kreisheimer, M., Schnelzer, M., and Tschense, A. (2006). Lung Cancer Risk Among German Male Uranium Miners: A Cohort Study, 1946–1998. *British Journal of Cancer,* 95 (9): 1280–1287.

[47]Guolo, A., and Brazzale, A. R. (2008). A Simulation-Based Comparison of Techniques to Correct for Measurement Error in Matched Case-Control Studies. *Statistics in Medicine,* 27: 3755–3775.

[48]Gustafson, P. (2004). Measurement Error and Misclassification in Statistics and Epidemiology - Impacts and Bayesian Adjustments. Chapman & Hall/CRC.

[49]Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.

[50]Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika,* 57 (1): 97–109.

[51]Hauptmann, M., Berhane, K., Langholz, B., and Lubin, J. (2001). Using Splines to Analyse Latency in the Colorado Plateau Uranium Miners Cohort. *Journal of Epidemiology and Biostatistics,* 6 (6): 417–424.

[52]Heid, I., Küchenhoff, H., Miles, J., Kreienbrock, L., and Wichmann, H. E. (2004). Two Dimensions of Measurement Error: Classical and Berkson Error in Residential Radon Exposure Assessment. *Journal of Exposure Analysis and Environmental Epidemiology,* 14: 365–377.

[53]Heid, I., Küchenhoff, H., Wellmann, J., Gerken, M., Kreienbrock, L., and Wichmann, H.-E. (2002). On the Potential of Measurement Error to Induce Differential Bias on Odds Ratio Estimates: An Example from Radon Epidemiology. *Statistics in Medicine,* 21: 3261–3278.

[54]Heid, I. (2002). Measurement Error in Exposure Assessment: An Error Model and Its Impact on Studies on Lung Cancer and Residential Radon Exposure in Germany (Thesis). PhD thesis, Ludwig-Maximilians-Universität.

[55]Heidenreich, W. F., Tomasek, L., Grosche, B., Leuraud, K., and Laurier, D. (2012). 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,* 51 (3): 263–275.

[56]Heidenreich, W. F., Luebeck, E. G., and Moolgavkar, S. H. (2004). Effects of Exposure Uncertainties in the TSCE Model and Application to the Colorado Miners Data. *Radiation Research,* 161 (1): 72–81.

[57]Hendry, D. J., (2014). Data Generation for the Cox Proportional Hazards Model with Time-Dependent Covariates: A Method for Medical Researchers. *Statistics in Medicine,* 33: 436–454.

[58]Higdon, R., and Schafer, D. W. (2001). Maximum Likelihood Computations for Regression with Measurement Error. *Computational Statistics & Data Analysis,* 35: 283–299.

[59]Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, 382–401.

[60]Hoffmann, S. (2017). Approche Hiérarchique Bayésienne Pour La Prise En Compte d'erreurs de Mesure d'exposition Chronique et à Faible Doses Aux Rayonnements Ionisants Dans L'estimation Du Risque de Cancers Radio-Induits: Application à Une Cohorte de Mineurs d'uranium." PhD thesis, Université Paris-Saclay (ComUE).

[61]Hoffmann, S., Guihenneuc, C., and Ancelet, S. (2018a). A Cautionary Comment on the Generation of Berkson Error in Epidemiological Studies. *Radiation and Environmental Biophysics,* 57 (2): 189–193.

[62]Hoffmann, S., Laurier, D., Rage, E., Guihenneuc, C., and Ancelet, S. (2018b). Shared and Unshared Exposure Measurement Error in Occupational Cohort Studies and Their Effects on Statistical Inference in Proportional Hazards Models. *PloS One,* 13 (2): e0190792.

[63]Hoffmann, S., Rage, E., Laurier, D., Laroche, P., Guihenneuc, C., and Ancelet, S. (2017). 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 (2): 196–209.

[64]Hu, P., Tsiatis, A. A., and Davidian, M. (1998). Estimating the Parameters in the Cox Model When Covariate Variables Are Measured with Error. *Biometrics,* 54: 1407–1419.

[65]Hughes, M. D. (1993). Regression Dilution in the Proportional Hazards Model. *Biometrics,* 49: 1056–1066.

[66]Janzen, D., and Saiedian, H. (2005). Test-Driven Development Concepts, Taxonomy, and Future Direction. *Computer,* 38 (9): 43–50.

[67]Jordan, M. I. (2004). Graphical Models. *Statistical Science,* 19 (1): 140–155.

[68]Kadane, J. B., and Wolfson, L. J. (1998). Experiences in Elicitation. *Journal of the Royal Statistical Society. Series D (the Statistician),* 47: 3–19.

[69]Keogh, R. H., Strawbridge, A. D., and White, I. R. (2012). Effects of Classical Exposure Measurement Error on the Shape of Exposure-Disease Associations. *Epidemiologic Methods,* 1 (1): 13–32.

[70]Kim, H.-M., Yasui, Y., and Burstyn, I. (2006). Attenuation in Risk Estimates in Logistic and Cox Proportional-Hazards Models Due to Group-Based Exposure Assessment Strategy. *Annals of Occupational Hygiene,* 50 (6): 623–635.

[71]Kreuzer, M., Sobotzki, C., Schnelzer, M., and Fenske, N. (2018). Factors Modifying the Radon-Related Lung Cancer Risk at Low Exposures and Exposure Rates Among German Uranium Miners. *Radiation Research* 189 (2): 165–176.

[72]Küchenhoff, H., Deffner, V., Aßenmacher, M., Neppl, H., Kaiser, C., Güthlin, D., et al. (2018). Ermittlung der Unsicherheiten der Strahlenexpositionsabschätzung in der Wismut-Kohorte - Teil I - Vorhaben 3616S12223. Ressortforschungsberichte zum Strahlenschutz. Bundesamt für Strahlenschutz (BfS).

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

[74]Küchenhoff, H., and Carroll, R. J. (1997). Segmented Regression with Errors in Predictors: Semi-Parametric and Parametric Methods. *Statistics in Medicine,* 16: 169–188.

[75]Laird, N., and Olivier, D. (1981). Covariance Analysis of Censored Survival Data Using Log-Linear Analysis Techniques. *Journal of the American Statistical Association,* 76 (374): 231–240.

[76]Landau, D. P., and Binder, K. (2014). *A Guide to Monte-Carlo Simulations in Statistical Physics*. Cambridge university press.

[77]Langholz, B., Thomas, D., Xiang, A., and Stram, D. (1999). Latency Analysis in Epidemiologic Studies of Occupational Exposures: Application to the Colorado Plateau Uranium Miners Cohort. *American Journal of Industrial Medicine,* 35 (3): 246–256.

[78]Lehmann, F. (2004). Job-Exposure-Matrix "Ionisierende Strahlung Im Uranerzbergbau Der Ehemaligen DDR.", Version 06/2004.

[79]Lehmann, F., Hambeck, L., Linkert, K.-H., Lutze, H., Meyer, H., Reiber, H., Reinisch, A., Renner, H.-J., Seifert, T., and Wolf, F. (1998). *Belastung Durch Ionisierende Strahlung Im Uranerzbergbau Der Ehemaligen DDR: Abschlußbericht Zu Einem Forschungsvorhaben*. Hauptverband der Gewerblichen Berufsgenossenschaften: Sankt Augustin.

[80]Lehmann, F., Lutze, H., Petter, W., Reiber, H., and Richter, S. (1994). Teilbericht: Strahlenexposition in den Aufbereitungsbetrieben und Beprobungszechen Der SAG/SDAG Wismut. Bergbau-Berufsgenossenschaft.

[81]Li, Y., and Lin, X. (2003). Functional Inference in Frailty Measurement Error Models for Clustered Survival Data Using the SIMEX Approach. *Journal of the American Statistical Association,* 98 (461).

[82]Liao, X., Zucker, D. M., Li, Y., and Spiegelman, D. (2011). Survival Analysis with Error-Prone Time-Varying Covariates: A Risk Set Calibration Approach. *Biometrics,* 67 (1): 50–58.

[83]Lubin, J. H., Boice Jr, J. D., Edling, C., Hornung, R. W., Howe, G., Kunz, E., Kusiak, R. A., Morrison, H. I., Radford, E. P., Samet, J. M., et al. (1995a). Radon-Exposed Underground Miners and Inverse Dose-Rate (Protraction Enhancement) Effects. *Health Physics,* 69 (4): 494–500.

[84]Lubin, J. H., Boice Jr, J. D., and Samet, J. M. (1995b). Errors in Exposure Assessment, Statistical Power and the Interpretation of Residential Radon Studies. *Radiation Research,* 144 (3): 329–341.

[85]Lubin, J. H., Wang, Z. Y., Wang, L. D., Boice Jr, J. D., Cui, H. X., Zhang, S. R., Conrath, S., Xia, Y., Shang, B., Cao, J. S., and Kleinerman, R. A. (2005). Adjusting Lung Cancer Risks for Temporal and Spatial Variations in Radon Concentration in Dwellings in Gansu Province, China. *Radiation Research,* 163 (5): 571–579.

[86]Marsh, J. W., Blanchardon, E., Gregoratto, D., Hofmann, W., Karcher, K., Nosske, D., and Tomášek, L. (2012). Dosimetric Calculations for Uranium Miners for Epidemiological Studies. *Radiation Protection Dosimetry,* 149 (4): 371–383.

[87]Martins, T. G., and Rue, H. (2012). Extending INLA to a Class of Near-Gaussian Latent Models. *arXiv Preprint arXiv:1210.1434*.

[88]Marušiaková, M., Gregor, Z., and Tomášek, L. (2011). A Review of Exposures to Radon, Long-Lived Radionuclides and External Gamma at the Czech Uranium Mine. *Radiation Protection Dosimetry,* 145 (2-3): 248–251.

[89]Masiuk, S., Shklyar, S., Kukush, A., Carroll, R., Kovgan, L., and Likhtarov, I. (2016). Estimation of Radiation Risk in Presence of Classical Additive and Berkson Multiplicative Errors in Exposure Doses. *Biostatistics,* 17 (3): 422–436.

[90]Messer, K., and Natarajan, L. (2008). Maximum Likelihood, Multiple Imputation and Regression Calibration for Measurement Error Adjustment. *Statistics in Medicine,* 27 (30): 6332–6350.

[91]Metropolis, N. (1987). The Beginning of the Monte Carlo Method. *Los Alamos Science Special Issue*, 125–130.

[92]Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics,* 21 (6): 1087–1092.

[93]Misumi, M., Kyoji F., Cologne, J. B., and Cullings, H. M. (2018). 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),* 67 (1): 275–289.

[94]Montez-Rath, M. E., Kapphahn, K., Mathur, M. B., Mitani, A. A., Hendry, D. J., and Desai, M. (2017). Guidelines for Generating Right-Censored Outcomes from a Cox Model Extended to Accommodate Time-Varying Covariates. *Journal of Modern Applied Statistical Methods,* 16 (1): 6.

[95]Muff, S., Ott, M., Braun, J., and Held, L. (2017). Bayesian two-component measurement error modelling for survival analysis using INLA — A case study on cardiovascular disease mortality in Switzerland. *Computational Statistics & Data Analysis,* 113: 177–193.

[96]Muff, S., Riebler, A., Held, L., Rue, H., and Saner, P. (2015). Bayesian Analysis of Measurement Error Models Using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 64 (2): 231–252.

[97]Murad, H., Kipnis, V., and Freedman, L. S. (2016). Estimating and Testing Interactions When Explanatory Variables Are Subject to Non-Classical Measurement Error. *Statistical Methods in Medical Research,* 25 (5): 1991–2013.

[98]Nolte, S. (2007). The Multiplicative Simulation Extrapolation Approach. Center for Quantitative Methods and Survey Research, University of Konstanz, Working Paper.

[99]Oakley, J. E., Daneshkhah, A., and O'Hagan, A. (2010). Nonparametric Prior Elicitation Using the Roulette Method. School of Mathematics; Statistics, University of Sheffield.

[100] Oh, E. J., Shepherd, B. E., Lumley, T., and Shaw, P. A. (2018). Considerations for Analysis of Time-to-Event Outcomes Measured with Error: Bias and Correction with SIMEX. Statistics in Medicine, 37 (8): 1276–1289.

[101] O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oackley, J. E., and Rakow, T. (2006). *Uncertain Judgements: Eliciting Experts' Probabilities*. John Wiley & Sons.

[102] Prentice, R. L. (1982). Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model. *Biometrika,* 69: 331–342.

[103] Richardson, D. B., Cole, S. R., Chu, H., and Langholz, B. (2011). Lagging Exposure Information in Cumulative Exposure-Response Analyses. *American Journal of Epidemiology*, 260.

[104] Richardson, S. (1996). Measurement Error. In *Markov Chain Monte Carlo in Practice*, edited by Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., 401–17. Chapman & Hall.

[105] Richardson, S., and Gilks, W. R. (1993a). A Bayesian Approach to Measurement Error Problems in Epidemiology Using Conditional Independence Models. *American Journal of Epidemiology,* 138 (6): 430–442.

[106] Richardson, S., and Gilks, W. R. (1993b). Conditional Independence Models for Epidemiological Studies with Covariate Measurement Error. *Statistics in Medicine,* 12: 1703–1722.

[107] Richardson, S., Leblond, L., Jaussent, I., and Green, P. J. (2002). Mixture Models in Measurement Error Problems, with Reference to Epidemiological Studies. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 165 (3): 549–566.

[108] Richter, S. (1994). Einfluß von Mängeln/Verstößen in Wetterführung und Strahlenschutz auf die Strahlenexposition unter Tage.

[109] Robert, C., and Casella, G. (2011). A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Statistical Science,* 26 (1): 102–115.

[110] Robert, C. P., and Casella, G. (2004). *Monte Carlo Statistical Methods - Second Edition*. Springer.

[111] Roberts, G. O., and Smith, A. F. (1994). Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms. *Stochastic Processes and Their Applications,* 49 (2): 207–216.

[112] Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 71 (2).

[113] Schafer, D. W., and Purdy, K. G. (1996). Likelihood Analysis for Errors-in-Variables Regression with Replicate Measurements. *Biometrika,* 83 (4): 813–824.

[114] Schiager, K. J., Borak, T. J., and Johnson, J. A. (1981). Radiation Monitoring for Uranium Miners: Evaluation and Optimization (Final Report) 9 Sept. 1979-9 Oct. 1981. Bureau of mines; Alara.

[115] Shonkwiler, R. W., and Mendivil, F. (2009). *Explorations in Monte Carlo Methods*. Springer Science & Business Media.

[116] Spiegelman, D., Logan, R., and Grove, D. (2011). Regression Calibration with Heteroscedastic Error Variance. *The International Journal of Biostatistics,* 7 (1).

[117] Stram, D., Langholz, B., Huberman, M., and Thomas, D. (1999). Correcting for Exposure Measurement Error in a Reanalysis of Lung Cancer Mortality for the Colorado Plateau Uranium Miners Cohort. *Health Physics,* 77 (3).

[118] Sylvestre, M.-P., and Abrahamowicz, M. (2008). Comparison of Algorithms to Generate Event Times Conditional on Time-Dependent Covariates. *Statistics in Medicine,* 27 (14): 2618–2634.

[119] Therneau, T., and Crowson, C. (2013). Using Time Dependent Covariates and Time Dependent Coefficients in the Cox Model. *The Survival Package (R Help Guide)*.

[120] Therneau, T., and Grambsch, P. (2000). *Modeling Survival Data, Extending the Cox Model*. Springer.

[121] Thiébaut, A. C. M., Freedman, L. S., Carroll, R. J., and Kipnis, V. (2007). Is It Necessary to Correct for Measurement Error in Nutritional Epidemiology? *Annals Internal Medicine,* 146 (1): 65–67.

[122] Tirmarche, M., Harrison, J., Laurier, D., Blanchardon, E., Paquet, F., and Marsh, J. (2012). Risk of Lung Cancer from Radon Exposure: Contribution of Recently Published Studies of Uranium Miners. *Annals of the ICRP,* 368–377 (3): 368–377.

[123] Torabi, M. (2013). Likelihood Inference in Generalized Linear Mixed Measurement Error Models. *Computational Statistics & Data Analysis*, 549–557.

[124] Vacquier, B., Rogel, A., Leuraud, K., Caer, S., Acker, A., and Laurier, D. (2008). Radon-Associated Lung Cancer Risk Among French Uranium Miners: Modifying Factors of the Exposure–Risk Relationship. *Radiation and Environmental Biophysics,* 48 (1): 1–9.

[125] van Walraven, C., Davis, D., Forster, A. J., and Wells, G. A. (2004). Time-Dependent Bias Was Common in Survival Analyses Published in Leading Clinical Journals. *Journal of Clinical Epidemiology,* 57 (7): 672–682.

[126] Volinsky, C. T., Madigan, D., Raftery, A. E., and Kronmal, R. A. (1997). Bayesian Model Averaging in Proportional Hazard Models: Assessing the Risk of a Stroke. *Applied Statistics,* 46 (4): 433–448.

[127] Walsh, L., Dufey, F., Tschense, A., Schnelzer, M., Grosche, B., and Kreuzer, M. (2010). Radon and the Risk of Cancer Mortality - Internal Poisson Models for the German Uranium Miners Cohort. *Health Physics,* 99 (3).

[128] Wang, C. Y., Hsu, L., Feng, Z. D., and Prentice, R. L. (1997). Regression Calibration in Failure Time Regression. *Biometrics,* 53 (131–145).

[129] Wismut GmbH. (1999). *Chronik Der Wismut*.

[130] Wolfson, L. J., and Bousquet, N. (2016). Elicitation. In *Wiley Statsref: Statistics Reference Online*, 1–11. John Wiley & Sons.

[131] Wolkewitz, M., Allignol, A., Harbarth, S., de Angelis, G., Schumacher, M., and Beyersmann, J. (2012). Time-Dependent Study Entries and Exposures in Cohort Studies Can Easily Be Sources of Different and Avoidable Types of Bias. *Journal of Clinical Epidemiology,* 65 (11): 1171–1180.

[132] Yi, G. Y., and Lawless, J. F. (2007). A Corrected Likelihood Method for the Proportional Hazards Model with Covariates Subject to Measurement Error. *Journal of Statistical Planning and Inference*, 1816–1828.

[133] Yi, G. Y., Ma, Y., Spiegelman, D., and Carroll, R. J. (2015). Functional and Structural Methods with Mixed Measurement Error and Misclassification in Covariates. *Journal of the American Statistical Association,* 110 (510): 681–696.

[134] Zettwoog, P. (1981). State-of-the-Art of the $alpha$ Individual Dosimetry in France. In *Radiation Hazards in Mining: Control, Measurements and Medical Aspects*, 4–9.

[135] Zhou, M. (2001). Understanding the Cox Regression Model with Time-Change Covariates. *The American Statistician,* 55 (2): 153–155.

A Appendix

A 1 Implementation of the algorithm

A 1.1 Overview

The main class of the algorithm is *MCMC* which serves as an interface for inference. It holds instances of the class `Parameter` and an instance of the class `LatentVariable`. When the sampling procedure is invoked, *MCMC* updates the state of all `Parameter` instances and the `LatentVariable` instance. The `LatentVariable` represents the (cumulative) exposure. The exposure of all workers in the cohort stems from different submodels represented through various uncertain factors. Therefore, the `LatentVariable` class contains a dictionary of instances of the class `UncertainFactor`. When *MCMC* updates the `LatentVariable`, the `LatentVariable` invokes updates for all its instances of the `UncertainFactor` class and calculates the cumulative exposure which is used in the updates of the parameters (invoked by *MCMC*). Each instance of the `UncertainFactor` represents its true unobserved value. It may hold some instances of the class `PriorParameterVector` which is a container of the class `PriorParameter`. A `PriorParameter` represents the prior on the distribution of an uncertain factor. For the case of $\mathcal{C}_{Rn}$, this is a collection of many instances represented by `PriorParameterVector`. This nested structure gives more flexibility in the estimation of $\mathcal{C}_{Rn}$ (we specify variable priors depending on the exposure year). An instance of the `UncertainFactor` class may hold instances of the classes `ErrorComponent` (representing a classical additive error), `MultClassicalErrorComponent` (representing a classical multiplicative error) and `BerksonErrorComponent` (representing a Berkson error which is assumed to be always multiplicative). We also implement a class `FixedParameter` which behaves like the `Parameter` class but skips updates. `FixedParameter` is used to fix the value of priors or of a parameter.

A 1.2 Updating the latent exposure

As we already described before the update for a parameter, we show here, how the update of the latent variable works using the `LatentVariable` class. Since the latent variable is a combination of different factors, we describe the update for one of these. When accounting for M2 in the Wismut cohort, the joint posterior distribution $[\theta|Y,X]$ can be expressed as:

$$
\begin{aligned}
[\theta, X|\cdot] = \quad & [\beta][\lambda][\alpha_\omega][\beta_\omega][\alpha_\gamma][\beta_\gamma][\alpha_\varphi][\beta_\varphi][\mu_{\mathcal{C}}][\sigma_{\mathcal{C}}] \times \\
& \prod_{i,t}[Y_i|\lambda,\beta,X_i^{cum}(t)] \times \\
& \prod_{i,t}[X_i(t)|\mathcal{C}_{R_n}(t,o),\varphi'(t,o,j),\gamma'(t,o),\omega'(t,o),{\tau_E}'(t),l(i,t,o,j)] \times \\
& \prod_{t,o}[\omega'(t,o)|\sigma^2_{\omega',B},\omega(p_t)]\prod_{p_t}[w(p_t)|\sigma^2_{\omega,c},\omega(p_t)]\prod_{p_t}[\omega(p_t)|\alpha_\omega,\beta_\omega] \times \\
& \prod_{t,o}[\gamma'(t,o)|\sigma^2_{\gamma',B},\gamma(p_t,o)]\prod_{p_t,o}[g(p_t,o)|\sigma^2_{\gamma,c},\gamma(p_t,o)]\prod_{p_t,o}[\gamma(p_t,o)|\alpha_\gamma,\beta_\gamma] \times \\
& \prod_{t,o,j}[\varphi'(t,o,j)|\sigma^2_{\varphi',B},\varphi(o,j)]\prod_{o,j}[f(o,j)|\sigma^2_{\varphi,c},\varphi(o,j)]\prod_{o,j}[\varphi(o,j)|\alpha_\varphi,\beta_\varphi] \times \\
& \prod_{t,o}[\mathrm{C}_{R_n}(t,o)|\sigma^2_{\mathcal{C},c},\mathcal{C}_{R_n}(t,o)]\prod_{t,o}[\mathcal{C}_{R_n}(t,o)|\mu_{\mathcal{C}},\sigma_{\mathcal{C}}]
\end{aligned}
$$

For a simple disease model in which the association between lung cancer mortality and radon exposure is modeled through a linear or log-linear function without effect modifying variables and in which the baseline hazard is modeled through a piecewise constant function with four different baseline hazards as a function of the age of the worker, the *MCMC* class contains a dictionary of

parameters, in which it holds five instances of the *Parameter* class (β, λ_1, λ_2, λ_3 and λ_4) and one instance of the LatentVariable class.
As described in Section *2.3*, the true exposure $X_i(t,o)$ of miner i in year t and object o in M2 depends on the product of the true working time factor in year t and object o $\omega'(t,o)$, the true equilibrium factor in year t and object o $\gamma'(t,o)$, the true activity weighting factor for activity j in year t and object o $\varphi'(t,o,j)$ and the true radon gas measurement in year t and object o $\mathcal{C}_{Rn}(t,o)$:

$$E(t,o,j) \quad = \mathcal{C}_{Rn}(t,o) \cdot 12 \cdot \gamma'(t,o) \cdot \omega'(t,o) \cdot \varphi'(t,o,j).$$

The *LatentVariable* class is responsible for the calculation of the true exposure of a miner working in year t in object o and activity j. To do so, it holds the four uncertain factors ω, γ, φ and $\mathcal{C}_{Rn}$. Each of these uncertain factors is an instance of the *UncertainFactor* class. The function *calculate_exposure()* in the *LatentVariable* class calculates the true exposure for each year t and each miner i. It does so by accessing the current values of the uncertain factors through the *get_values()* function which is implemented in the *UncertainFactor* class for each uncertain factor ω, γ, φ and $\mathcal{C}_{Rn}$ and by multiplying the values by 12.
Each instance of *UncertainFactor* contains at least one instance of the class *ErrorComponent* or *MultClassicalErrorComponent* which represents the classical measurement error component of the *UncertainFactor* instance. Depending on the error structure, the *UncertainFactor* may also hold a *BerksonErrorComponent* representing the Berkson error component.
In order to avoid too much repetition in the presentation, we will describe the implementation and the functioning of the algorithm while focusing on the working time factor $\omega'(t,o)$ in the following. For this uncertain factor, the Berkson and classical measurement error component are given by

$$\begin{aligned} w(p_t) \quad &= \omega(p_t) \cdot U_{\omega,c}(p_t) \\ \omega'(t,o) \quad &= \omega(p_t) \cdot U_{\omega',B}(t,o). \end{aligned}$$

The algorithm proposes new values for the classical and for the Berkson errors jointly. Therefore, for the update of $\omega(p_t)$ and $\omega'(t,o)$, we have to look at the joint posterior which is given by:

$$\begin{aligned} [w(p_t), \omega'(t,o)|\cdot] \propto \quad & \prod_{i,t} [\, Y_i|\lambda, \beta, X_i^{cum}(t)] \times \\ & \prod_{t,o} [\, \omega'(t,o)|\sigma^2_{\omega',B}, \omega(p_t)] \times \\ & \prod_{p_t} [\, w(p_t)|\sigma^2_{\omega,c}, \omega(p_t)] \times \\ & \prod_{p_t} [\, \omega(p_t)|\alpha_\omega, \beta_\omega] \end{aligned}$$

These are the only terms that either depend on $\omega'(t,o)$ or $\omega(p_t)$.

Since we have the combination of classical and Berkson errors over different uncertain factors representing the latent exposure, the implementation of the update of the latent variable faces more challenges: The classical error and the Berkson error may affect the exposure on different dimensions. For instance, the classical error on the factor ω depends only on p_t while the Berkson error depends on t and o. To solve this problem, we suggest a 2-step procedure. We show this procedure in the general update flow below for ω after showing the update flow for one parameter.

A 1.3 General update flow of the algorithm for one iteration

We showed how the update of a parameter within an MCMC algorithm and for an uncertain factor as part of the latent exposure variable is made.
In the following, we will present an update for an arbitrary unknown parameter of the measurement model and afterwards an update for ω will be described to get more insight in the algorithm. Note that all ratios are calculated on the log scale and get transformed back to the normal scale right

before the final acceptance rate calculation. This ensures numerical stability and speeds up some computations due to the simplifications in the used densities.
The algorithm iterates over all parameters and updates each of them. Afterwards, the constituent parts of the latent variable are getting updated.

Update scheme for a parameter

The update for a parameter is conducted in a separate Python module *`update.py`*. Note that for a parameter update the latent variable is fixed (i.e. as described before, the update can be done given the values of the latent variable). The update scheme for the parameter is rather straightforward and it works as follows:

1. The *MCMC* class requires the current state of the chain (all parameters and the latent variable) and passes it to the update function within the *`update.py`* module.
2. A new value is proposed given the current value. All other values stay the same.
3. Using the proposed value of the parameter to be updated and the current values of the other parameters and the latent variable, the acceptance rate ρ is calculated as described in Section 4.6.1.
4. Accept the proposed value as new state with probability ρ.
5. Increment the iterator from t to t+1 for that parameter and store the updated information.

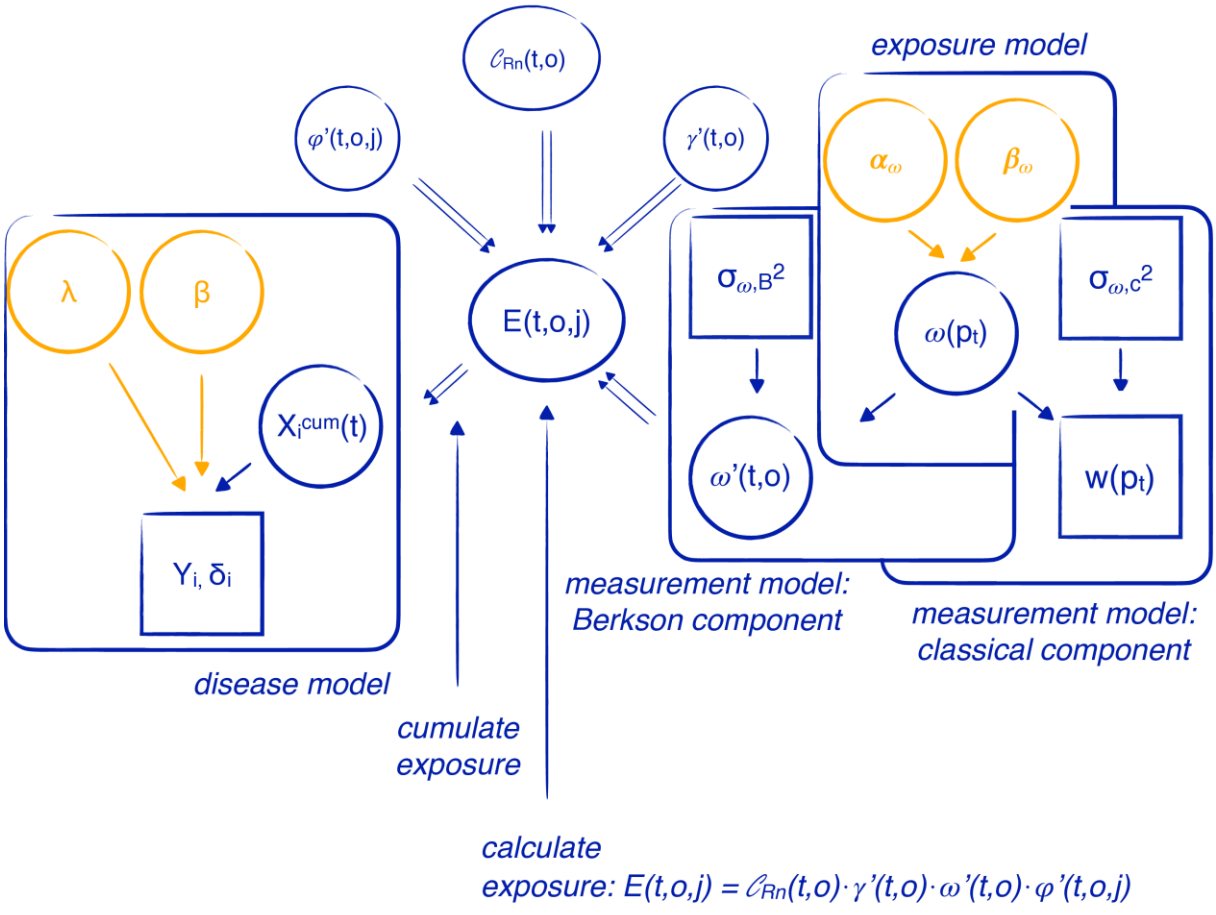

Figure A.1: The disease model, Berkson and classical measurement model of ω and the exposure model for the update of ω

Update scheme for ω

The update for each of the uncertain factors follows in general the same scheme as described before. However, since the latent exposure is a combination of different uncertain factors which have a complex error structure, the update of each of them is more complex. The update is exemplary illustrated for ω. Figure *A.1* shows the focus on the relevant parts of the full DAG. If ω is updated, the other uncertain factors can be seen as fixed. The same is true for the parameters β and λ. The update scheme has to include two measurement models: One for the Berkson and one for the classical component. The full update scheme is given as follows:

1. The true mean values of ω depend on p_t. Therefore, the algorithm uses as many different mean values as the number of different time periods. Given the current state, a candidate for new mean values is proposed (for each p_t) and the actual classical measurement errors are calculated (one for each p_t). Depending on a multiplicative or additive error structure, the error is calculated as $U = Z/X$ or $U = Z - X$. For ω the error is therefore calculated as $U_{\omega,c}(p_t) = \omega(p_t)/\mathrm{w}(p_t)$.
2. Given the calculated errors and the proposed value, the proposal ratio, the measurement ratio and the exposure ratio can be calculated.
3. After the classical error is calculated, the Berkson error must be taken into account. Since the Berkson error affects the uncertain factor in another dimension, it is necessary to expand the uncertain factor to the correct dimension. To do that, a mapping matrix is used. The current true mean values are multiplied by this matrix to map it to the right dimension. In the case of ω, the dimension is defined over all time points and objects. Afterwards the Berkson errors $U_{\omega\prime,B}$ are proposed and multiplied on the resulting values with the correct dimension yielding $\omega'(t,o)$. To understand this procedure, see the following example:

$$\omega'(t,o) = \omega(p_t) \cdot U_{\omega\prime,B}(t,o) \text{ with}$$

$$\begin{pmatrix} \omega'(1955,1) \\ \omega'(1955,2) \\ \omega'(1956,1) \\ \omega'(1957,1) \\ \omega'(1958,1) \\ \omega'(1959,1) \\ \omega'(1959,2) \\ \omega'(1967,1) \\ \omega'(1968,1) \\ \omega'(1968,2) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \omega(1) \\ \omega(2) \\ \omega(3) \end{pmatrix} \circ \begin{pmatrix} U_{\omega\prime,B}(1955,1) \\ U_{\omega\prime,B}(1955,2) \\ U_{\omega\prime,B}(1956,1) \\ U_{\omega\prime,B}(1957,1) \\ U_{\omega\prime,B}(1958,1) \\ U_{\omega\prime,B}(1959,1) \\ U_{\omega\prime,B}(1959,2) \\ U_{\omega\prime,B}(1967,1) \\ U_{\omega\prime,B}(1968,1) \\ U_{\omega\prime,B}(1968,2) \end{pmatrix}$$

   Here the dimension of $\omega(p_t)$ is three (3 different periods). The dimension for the Berkson errors affects the mean values on the time scale and objects. In this example, the mapping matrix has therefore a dimension of 10×3. Afterwards the Berkson errors are multiplied element-wise on the new dimension resulting in proposed deviations of $\omega'(t,o)_{\text{cand}}$.
4. After proposing Berkson errors, the measurement ratio and the proposal ratio for these values can be calculated.

5. Given new values with the Berkson error, it is again necessary to map them to the full dimension of the data set. This is again done by a mapping matrix in the same manner as before but with another dimension (if the model is specified without a Berkson error, the multiplication of the first mapping matrix is not necessary). Since the true exposure is defined as $X_i(t,o) = \mathcal{C}_{Rn}(t,o) \cdot 12 \cdot \gamma'(t,o) \cdot \omega'(t,o) \cdot \varphi'(t,o,j) \cdot l(i,t,o,j)$, the new proposed latent exposure can be calculated as $X_i(t,o)_{\text{cand}} = X_i(t,o)_{\text{t}} / \omega'(t,o)_{\text{t}} \cdot \omega'(t,o)_{\text{cand}}$.
6. Afterwards, the new latent exposure values are cumulated. This is achieved by using again a sparse matrix with a 1 on triangular blocks to sum up the exposures received by each of the workers. As an illustration, see the following short example:

$$\begin{pmatrix} X_1(1) \\ X_1(1)+X_1(2) \\ X_1(1)+X_1(2)+X_1(3) \\ X_1(1)+X_1(2)+X_1(3)+X_1(4) \\ X_2(1) \\ X_2(1)+X_2(2) \\ X_3(1) \\ X_3(1)+X_3(2) \\ X_3(1)+X_3(2)+X_3(3) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \\ X_{14} \\ X_{21} \\ X_{22} \\ X_{31} \\ X_{32} \\ X_{33} \end{pmatrix}$$

7. Given the new proposed cumulative exposure values, it is now possible to calculate the ratio of the disease model.
8. To get the final acceptance probability, the six ratios of the different models must be summed up and exponentiated:
   - The proposal, measurement and exposure ratio from the classical error component (calculated in step 2)
   - The proposal and measurement ratio from the Berkson component (calculated in step 4)
   - The ratio of the disease model
9. The proposed latent exposure (through $\omega'(t,o)_{\text{cand}}$) and the proposed errors are accepted as new state of the system according to the acceptance probability calculated in step 8, otherwise the old state is carried forward.
10. In a last step, the iterators are incremented from t to t+1 and the update information is stored.

A 2 Generated survival times in the simulation study

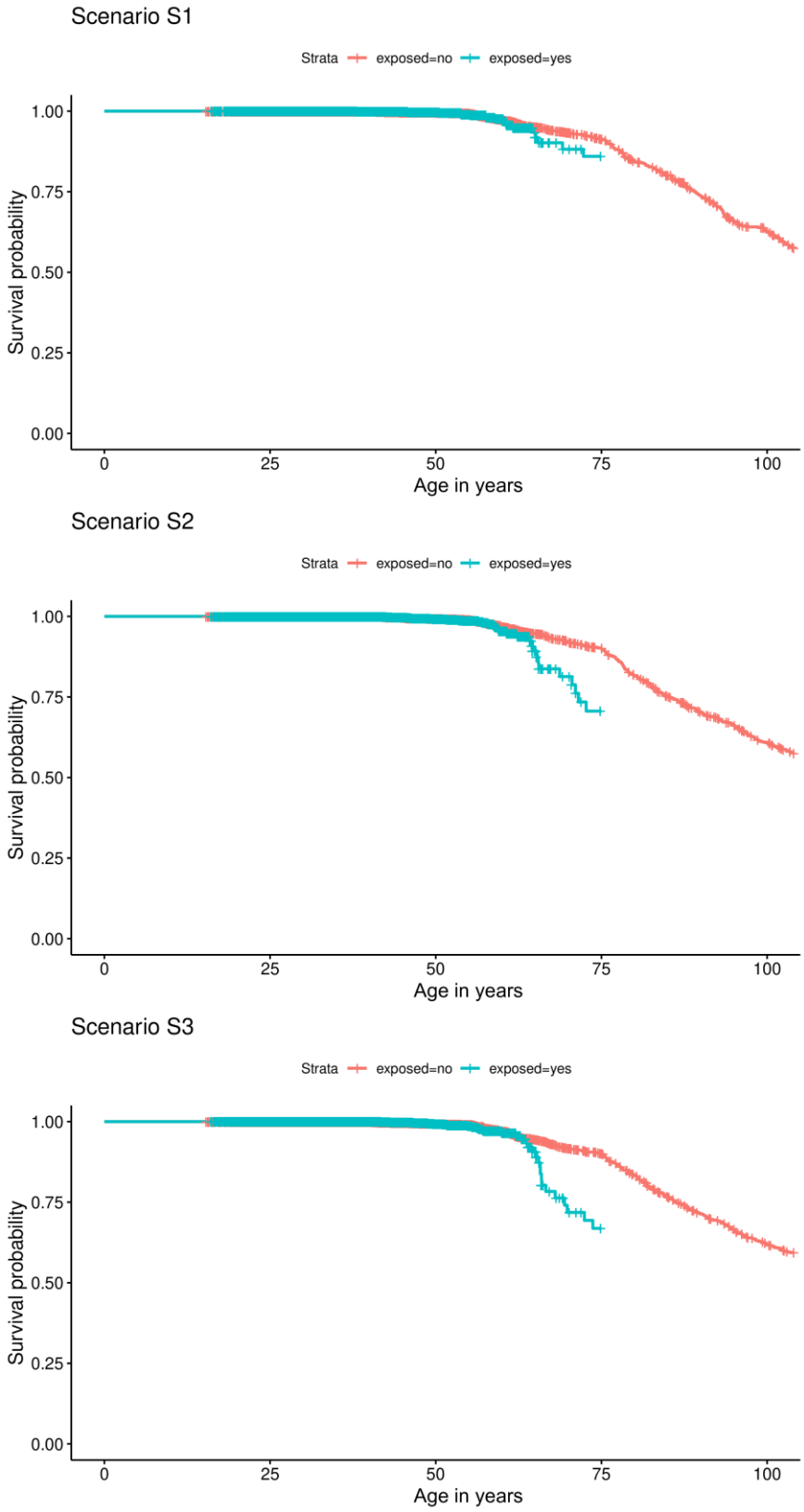

Figure A.2: Kaplan Meier curves for non-exposed miners with 0 WLM (red) and exposed miners (blue) for one simulation data set generated according to scenarios S1, S2 or S3, respectively.

A 3 Estimated 95% credible intervals for the proposed Bayesian hierarchical approach in the simulation study

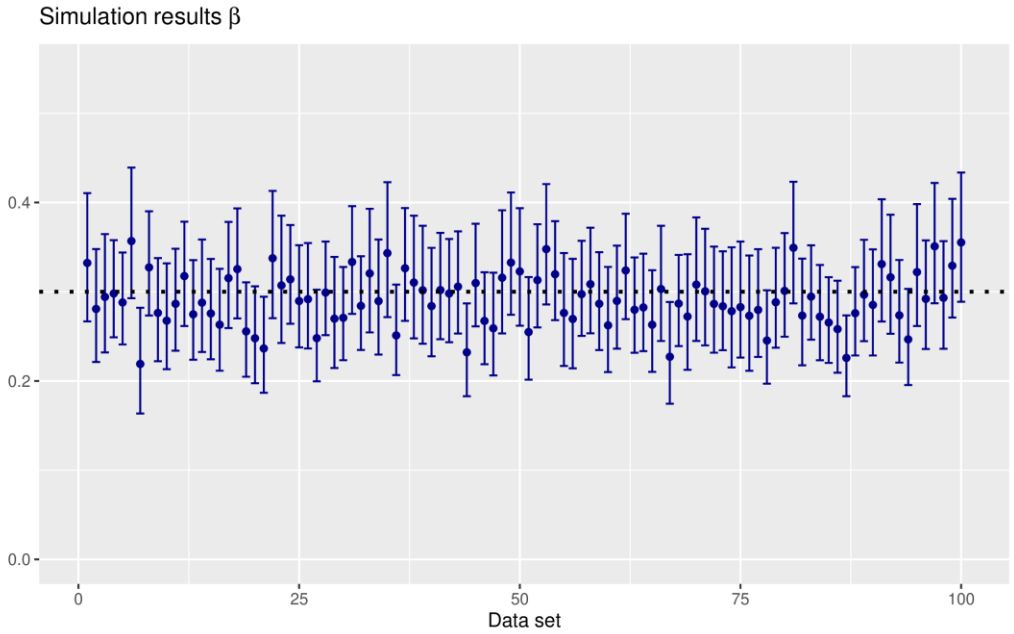

Figure A.3: Estimated means and 95% credible intervals on 100 data sets for the parameter β of the Cox model using the Bayesian hierarchical approach for simulation scenario S1. The dotted horizontal line is the true underlying value of β.

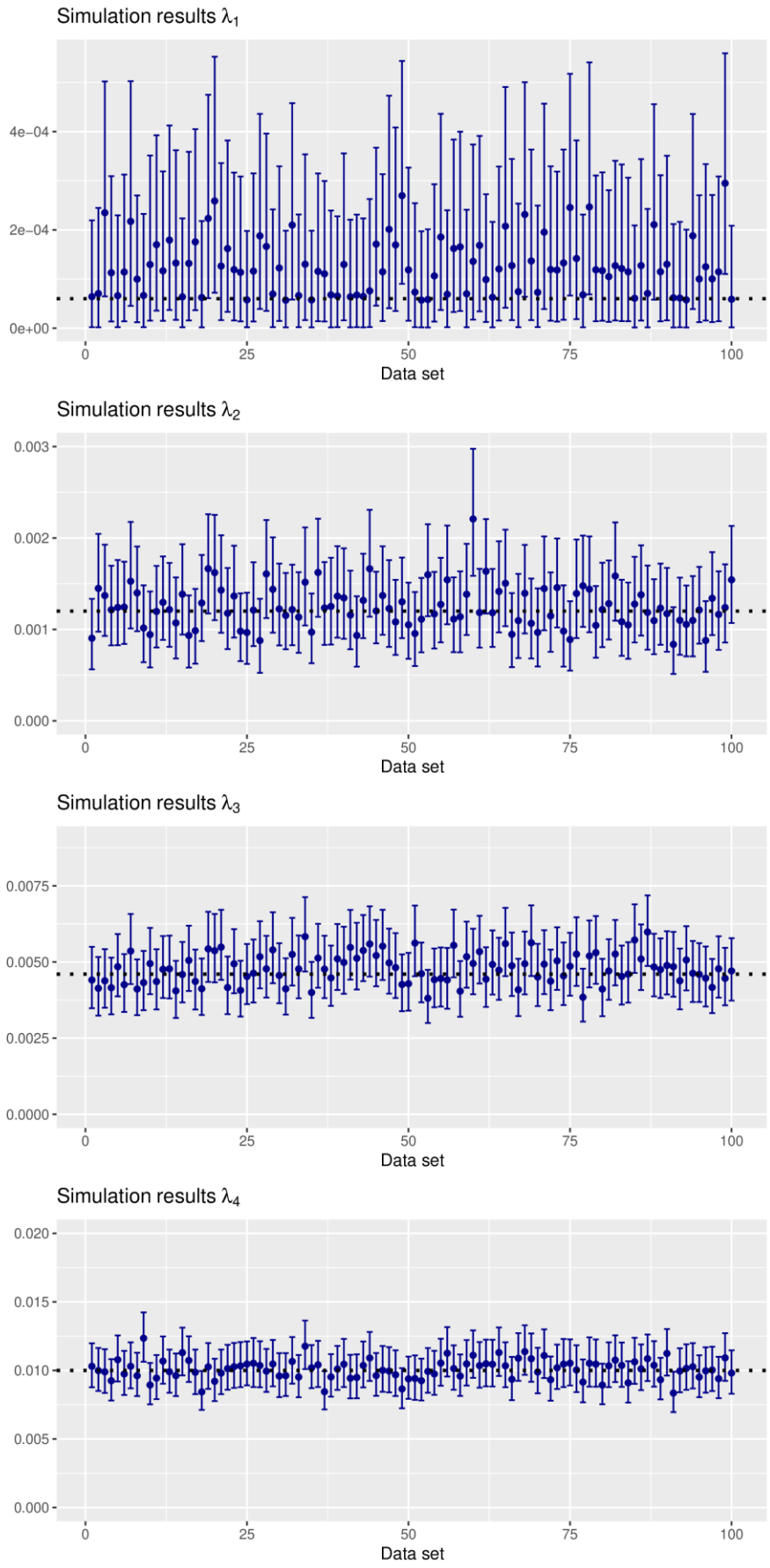

Figure A.4: Estimated means and 95% credible intervals on 100 data sets for the baseline hazards ($\lambda_k, k = 1,2,3,4$) of the Cox model using the Bayesian hierarchical approach for simulation scenario S1. The dotted horizontal line is the true underlying value.