From
the Institute of Medical Information Processing, Biometry and Epidemiology,
Ludwig-Maximilians-Universität München, Germany
Chair of Epidemiology: Prof. Dr. Dr. H.-E. Wichmann
and
the Institute of Epidemiology,
GSF-National Research Centre of Environment and Health, Neuherberg, Germany
Director: Prof. Dr. Dr. H.-E. Wichmann

Measurement Error in Exposure Assessment: An Error Model and its Impact on Studies on Lung Cancer and Residential Radon Exposure in Germany

Thesis
in fulfilment of the requirements for the degree
'Doktor der Humanbiologie (Dr. rer. biol. hum.)'
submitted to the Medical Faculty
of the Ludwig-Maximilians-Universität at Munich

By
Iris M. Heid
from
Munich, Germany
2002

With approval of the medical faculty
of the Ludwig-Maximilians-Universität München

First reviewer:	Prof. Dr. Dr. H.-E. Wichmann
Second reviewer:	Prof. Dr. A. M. Kellerer
	Prof. Dr. D. Nowak
Statistical supervision:	PD Dr. H. Küchenhoff
Dekan:	Prof. Dr. med. Dr. h. c. K. Peter
Date of oral examination:	09.10.2002

1 INTRODUCTION 7

1.1 RISK FACTORS OF LUNG CANCER 7

1.1.1 Smoking 7

1.1.2 Radon 8

1.1.3 Other risk factors 9

1.2 STUDIES ON LUNG CANCER AND RESIDENTIAL RADON EXPOSURE 10

1.2.1 The case-control study design 10

1.2.2 Multiple logistic regression and the control for confounder 11

1.2.3 Three generations of radon studies and the role of measurement error 11

1.3 THE GERMAN RADON STUDIES 13

1.3.1 Study subjects and exposure assessment 13

1.3.2 Descriptive statistics and naive odds ratio estimates 14

1.3.3 Contrasting the German West study region with West-High 16

1.3.4 Contrasting the German West and the East study 17

1.4 ACCOUNTING FOR MEASUREMENT ERROR IN EXPOSURES 17

1.5 OBJECTIVES 19

2 MATERIAL AND METHODS 21

2.1 MEASUREMENT ERROR CORRECTION METHODS 21

2.1.1 The tool box 22

2.1.1.1 Model for the distribution of the true predictors 22

2.1.1.2 The true exposure-disease model 24

2.1.1.3 Error model 25

2.1.2 Regressing the true predictors on the observed predictors 29

2.1.2.1 Ignoring the association between predictors 30

2.1.2.2 Accounting for the association between predictors 30

2.1.3 Regression calibration method to account for errors in two predictors 32

2.1.3.1 Ignoring the association between predictors 33

2.1.3.2 Accounting for the association between predictors 34

2.1.3.3 Computation of beta-coefficients and confidence intervals 35

2.1.4 Approximate Likelihood Method *35*

2.1.4.1 Ignoring the association between predictors 36

2.1.4.2 Accounting for the association between predictors 36

2.1.4.3 Computation of beta-coefficient estimates and confidence intervals 37

2.1.5 Sensitivity analysis *38*

2.2 REPLICATION DATA 39

2.2.1 Data *39*

2.2.1.1 Bedroom and living room measurements 39

2.2.1.2 Year-by-year replicates 40

2.2.1.3 Data from laboratory intercomparison study 40

2.2.2 Method of analysis *41*

2.2.2.1 Bedroom and living room measurements 41

2.2.2.2 Year-by-year replicates 42

2.2.2.3 Data from laboratory intercomparison study 43

3 RESULTS 44

3.1 THE ERROR MODEL IN THE GERMAN RADON STUDIES 44

3.1.1 Predictor variable of interest *44*

3.1.1.1 Concentration, exposure or dose 44

3.1.1.2 Radon concentration, radon exposure and alpha dose 46

3.1.1.3 Number of smoked cigarettes, packyears and lung dose of inhaled carcinogens 47

3.1.1.4 The operationally defined true predictor 48

3.1.2 Sources of error *50*

3.1.2.1 Errors in the radon variable 50

3.1.2.2 Errors in the smoking variable 55

3.1.3 Error structure *57*

3.1.3.1 Classical versus Berkson error 57

3.1.3.2 Multiplicativity versus additivity of error 61

3.1.3.3 Random versus systematic error 62

3.1.3.4 Homoscedastic versus heteroscedastic error 62

3.1.4 Error size *63*

3.1.4.1 Some measures of error size 63

3.1.4.2 Proportion of error variance to the variance of the observed predictor 64
3.1.4.3 Range of observed predictor values given a certain true value 66
3.1.5 Information from replicate data 70
3.1.5.1 Bedroom and living room measurements 70
3.1.5.2 Year-by-year replicates 71
3.1.5.3 Intercomparison data 74
3.2 INTERPRETING THE CORRECTION METHODS 75
3.2.1 Errors in the predictor of primary interest without covariate 76
3.2.1.1 Impact on the effect estimate of primary interest 76
3.2.1.2 Dependence of the bias on predictor distribution parameters 80
3.2.2 Including an error-free covariate 83
3.2.2.1 Impact on bias from error in the primary variable 83
3.2.2.2 Dependence of the bias on the correlation 84
3.2.3 Including an error-prone covariate 85
3.2.3.1 Impact on the effect estimate of primary interest 85
3.2.3.2 Dependence of the bias on the correlation 86
3.2.4 Confidence intervals 88
3.2.5 Approximate likelihood method by Reeves et al. 90
3.3 IMPACT OF ERRORS IN RISK FACTORS IN THE GERMAN RADON STUDIES 92
3.3.1 Correcting for errors in radon exposure and packyears (multiplicative classical error) 93
3.3.1.1 Correcting the correlation for multiplicative classical error 93
3.3.1.2 Correcting the OR estimates for multiplicative classical error: Full data 94
3.3.1.3 Correcting for multiplicative classical error: Excluding one outlier 99
3.3.2 Sensitivity on the choice of the error model (additive classical error) 101
3.3.2.1 Correcting for additive classical error: Full data 101
3.3.2.2 Correcting for additive classical error: Excluding one outlier 105
3.3.3 Accounting for the use of external exposure as a surrogate for lung dose (Berkson error) 107
3.3.3.1 Correcting for multiplicative Berkson error in radon exposure: Full data 107
3.3.3.2 Correcting for multiplicative Berkson error in radon exposure: Excluding one outlier 109
3.3.4 Comparison to correction with the approximate likelihood method 111
3.3.4.1 Correcting for multiplicative classical error 111

3.3.4.2 Correcting for multiplicative Berkson error 113

4 DISCUSSION 115

4.1 Error in assessing radon exposure (multiplicative classical error) 116

4.2 Reversal of effect in the West study 117

4.3 Errors in assessing packyears (multiplicative classical error) 119

4.4 Impact of the correlation between radon exposure and smoking 120

4.5 Residual confounding 122

4.6 Comparing West and West-High 124

4.7 Concordance between West and East 126

4.8 The sensitivity on the choice of a multiplicative versus additive error 127

4.9 Error from using external exposure as a surrogate for lung dose (multiplicative Berkson error) or „Does the lung dose really matter?“ 127

4.10 Correction methods 129

4.11 The error model 130

4.12 Other model assumptions 136

4.13 Notes of caution 137

5 SUMMARY 139

6 ZUSAMMENFASSUNG 140

7 A GUIDE TO DEFINITIONS AND ABBREVIATIONS 141

7.1 List of definitions 141

7.2 List of abbreviations 142

8 REFERENCES 143

9 ACKNOWLEDGEMENTS 150

1 INTRODUCTION

Cancer is the second leading cause of death in North-America and Europe. Lung cancer is the most frequent neoplasm among men in the U.S.A and in Europe and rapidly increasing among women. Lung cancer incidence has grown since the second world war in the industrialised nations, which is mainly attributed to the growing prevalence of smoking, the most prominent risk factor. Lung cancer incidence does not only depend upon gender, but also increases with age, and occurs more frequently in industrialised areas when compared to more rural areas (Doll et al. 1981, Hammond and Horn, 1958).

1.1 Risk factors of lung cancer

1.1.1 Smoking

Cigarette smoke contains more than 3800 components (gases and particles) with numerous carcinogens among them (IARC 1986) for example benzo[a]pyrene, arsenic, nickel and the radioactive Polonium, ^{210}Po. In addition, polycyclic aromatic hydrocarbons and nitrosamines contained in cigarette smoke are regarded as promoters of cancer development.

The association between lung cancer and smoking has been first described by Doll and Hill (1952). It is now generally accepted that lung cancer risk increases with the number of cigarettes smoked per day, age at start of smoking and the years of smoking, as well as depth of inhalation and the tar concentration in the cigarettes smoke. About 80% of the lung cancer cases in men and about 40% in women are attributed to smoking (Hammon et al. 1980). A cumulative measure for smoking is *packyear*, the product of the number of daily smoked cigarettes and the number of years of smoking divided by 20 (assuming 20 cigarettes per pack). Currently smoking men with more than 40 packyears (e.g. smoking 2 packs a day for at least 20 years) increase their lung cancer risk by 50 when compared to a male never-smoker (Wichmann et al., 1999). Smoking pipes or cigars is also associated with an increased lung cancer risk, but is, however, way below the risk of smoking cigarettes.

1.1.2 Radon

For uranium miners, studies have shown an increased risk of lung cancer, which is due to the radioactive alpha decay of radon and its progeny, ^{218}Po, ^{214}Po, and ^{210}Po. In fact, the "Schneeberger Lungenkrankheit", a lethal "lung disease" found particularly in the mining area around Schneeberg was described as early as the 16th century (Agricola, 1928) and is thus one of the longest known occupational disease. Animal experiments give further proof of the carcinogenic effect of radon and its progeny (e.g. IARC 1988, EPA 1987, 1992, NAS 1988, 1994, UNSCEAR 1977, 1982, 1986, ICRP 1987, 1990, 1993, SSK 1985, 1988, 1992).

Since the largest proportion of radon is exhaled immediately before the alpha decay (half-life of 3,82 days), the ionising carcinogenic effect stems mainly from the deposition of the short-living radon progeny, ^{218}Po (3 minutes) and ^{214}Po (164 microseconds) in the bronchial system. The alpha dose resulting from a given exposure to radon depends upon the equilibrium between radon and the short-living progeny, the fraction of the particle-bound progeny, aerosols in the inhaled air, respiratory characteristics (frequency, volume), bronchial geometry, mucociliar clearance, thickness of airway mucus and the location of the target cells (e.g. Jacobi et al. 1980, 1964).

Radon does not only appear in uranium mining. It is an ubiquitous gas evaporating from the ground due to uranium containing rock such as granite. The gas can accumulate in homes giving rise to concentrations potentially high enough to increase the lung cancer risk of the general population as was initially suggested by extrapolation of radon concentrations experienced by miners to indoor concentrations (Lubin et al. 1994). Although the lung cancer risk associated with radon levels found in homes is small, it is a public health issue, which concerns the general population. It is estimated that about 12% of the lung cancer cases in men and about 4% in women are attributed to residential radon exposure in Western Germany (Steindorf et al. 1995). A measure for radon and radon progeny in the air is the activity concentration, that is the number of alpha decays per second given in *Becquerel* (Bq) per cubic metre. Individuals subjected to over 140 Bq/m³ at their residence have a 1.3-fold risk of lung cancer when compared to someone subjected to 0-5 Bq/m³ (Wichmann et al., 1999).

The magnitude of radon concentrations is influenced by a variety of factors: Outdoor radon concentrations depend on the radioactive decay rate, diffusion, convection, wind, air current

and precipitation, temperature and compression (Gesell, 1983). Higher concentrations are usually found during the winter due to the more commonly encountered inversion. The seasonal differences vary about a factor of 2 to 4. Daily differences vary about a factor of 2 to 5 with higher concentrations during the late night and early morning, which is due to atmospheric stability and convection depending on daytime (Michel, 1987). Indoor radon concentrations depend on the geology of the ground below the building (content of radon precursors, porosity, density, humidity, permeability), on the vicinity and isolation against the ground (floor level, existence of a basement), on the difference of pressure between ground and building (due to temperature, compression) and thus on heating system, on exhalation of radon from building material, on the exchange of fresh air with indoor air cumulated with radon and its progeny (isolation of windows, ventilation) (Gerken et al., 2000, Gunby et al. 1993).

1.1.3 Other risk factors

Occupational risk factors associated with increased lung cancer risk are: asbestos, arsenic, beryllium, chromium and nickel, coal tar and soot, diesel motor emissions, and polycyclic aromatic hydrocarbons (Brüske-Hohlfeld et al., 2000, Boffetta et al., 1995). Environmental tobacco smoke may also be considered as an occupational risk factor (Kreuzer et al., 2001). Several of these factors are associated with a high lung cancer risk, but the fraction of exposed individuals is usually small. The estimated proportions of lung cancer attributable to occupational exposures range from 1% to 40% depending on the prevalence of the exposure in the local community (Vineis and Simonato, 1991). Other suspected or recognised risk factors of lung cancer include indoor (Pershagen et al., 1994, Mumford et al. 1987) and outdoor air pollution (Jedrychowski et al., 1990, Jöckel et al., 1992), and a family history of lung cancer (Ooi et al., 1986, Kreuzer et al., 1998).

1.2 Studies on lung cancer and residential radon exposure

1.2.1 The case-control study design

Ecologic studies compare the frequency of the risk factor and disease between different populations based on aggregated data instead of using individual data. They are easily designed, quickly conducted and apt to generate hypotheses. However, this study design is not adequate to provide evidence in favour of an association between a risk factor and the disease, because it cannot be identified if the disease and the risk factor occur simultaneously, neither can the impact of two or more factors be distinguished and thus a confounder can not be controlled for (Kleinbaum et al. 1982, Rothman 1986, Wichmann and Kreienbrock 1992) [Note that a *confounder* is defined as a factor that is associated with both the risk factor of interest (here: radon) and the disease (here: lung cancer); further, a confounder must be in the causal pathway between the risk factor of interest and the disease (Rothman, 1998, p.120)].

The concept of a case-control study using individual data is the most appealing for this study objective, particularly due to the fact that radon exposure is a "common" exposure; in fact everybody is exposed to some extent (Breslow, 1980). Hereby, a pre-defined number of lung cancer *cases* are recruited from hospitals or general practices, and suited *controls* are recruited for example via population registry or telephone registry (*population controls*), or alternatively from hospitals (*hospital controls*). In a basic analysis, the group of cases is compared with the group of controls with regard to the different prevalence of the risk factor (if binary) or different mean level of the risk factor (if continuous).

It is important that information on a potential confounder (here: age, gender, smoking) is gathered on an individual basis. There are three methods to deal with confounders. The analysis can be *stratified* for groups of individuals, which are similar with respect to the confounder. Another method attempts to control for the confounder by study design: The controls can be selected to be similar with regard to the confounding variable (*matching*); a certain number of controls is chosen for ach case (*individual matching*), or a group of controls is chosen to equals a group of cases with regard to the distribution of the confounder (*frequency matching*). If more than three factors are involved, particularly if some of them are continuous, a multiple regression analysis is to be preferred to stratification or matching.

1.2.2 Multiple logistic regression and the control for confounder

Multiple regression analysis is an analysis based on a mathematical model (exposure-disease model) relating the disease outcome (dependent variable) and several risk factors (independent variables). It allows the estimation of the contribution of the risk factors to the disease outcome.

Stating the exposure-disease model is the most prominent assumption made by this type of analysis. Most commonly, the exposure-disease relationship is assumed to be linear. Since the (multiple) *logistic regression* model is linear for small to moderate levels of exposure, the statistical properties of this model are good, and the regression coefficients are interpretable as the logarithm of the relative disease risk from one risk factor, this is the mostly applied exposure-disease model in epidemiology.

An identified confounder, for which data is collected for each individual, can be controlled for by including it into the model as an independent variable. If, however, the confounder is included into the model in a way that does not reflect the truth (e.g. including a variable linearly, whereas the impact on the disease depends quadratically upon the factor), if the confounder is assessed with error, or if other unidentified confounders exist, residual confounding may distort the estimate of the disease risk of primary interest (here: lung cancer risk due to radon exposure).

1.2.3 Three generations of radon studies and the role of measurement error

Since radon has been acknowledged as an important public health issue due to the severity of disease, lung cancer, and the ubiquitous appearance of this risk factor, several epidemiological studies, not only case-control, but also cohort studies, have been conducted throughout the world investigating the magnitude of the lung cancer risk from radon. These studies can be classified into three generations (Pershagen, 1993, Gerken et al., 1997):

The first generation includes studies from Sweden, Canada and the U.S.A., where the type of house, distance of living areas to the ground, or the geology of the residence's ground mainly serve as a substitute for individual radon concentration measurements (Axelson et al., 1979,

Simpson et al., 1983, Edling et al., 1984, Damber et al., 1987, Less et al., 1987, Svensson et al., 1987, Axelson et al., 1988, Klotz et al., 1989). The results of these studies range from a increased lung cancer risk for some house types (stone houses in Sweden, Axelson et al., 1979) or dwellings close to the ground (Svensson et al., 1987) associated with higher radon levels to no risk at all (Damber et al., 1987). These studies are limited due the use of crude surrogates for elevated radon concentrations in the homes, and a need of radon measurements in the homes of each study participant has been emphasised.

This has motivated the second generation of radon studies in China, U.S.A, Finland and Sweden (Blot et al. 1990, Schoenberg et al., 1990, Ruosteenoja, 1990, Pershagen et al., 1992), with individual radon measurements, the latter three of which found an increased risk, whereas the first one (Blot et al., 1990) found none. The main points of criticism regarding these studies have been the small sample size (200-400 cases) and the potential of measurement error in the radon exposure assessment. Bäverstam et al., 1991, have elucidated various sources of error, such as the error from pure between-measurement-variability given by the measuring process, the extrapolation of the radon concentration of the measured year to prior years, error from seasonal variations, error from missing measurements or from varying occupancy of the homes, from exposure outside of the home, as well as error from measuring radon instead of radon progeny and from using environmental radon exposure as a surrogate for the effective lung dose.

The third generation of radon studies in Europe (Pershagen et al., 1994, Auvinen et al., 1996, Darby et al., 1998, Wichmann et al., 1998, 1999, Kreienbrock et al., 2001) and North-America (Schoenberg et al, 1992, Létourneau et al.. 1994, Alavanja et al., 1994, 1999, Sandler et al, 1999, Field et al., 2000) address these issues by increased study size (500 to up to 1500 cases) and improved exposure assessment (e.g. assessing occupancy time, measuring radon concentrations not only in the homes inhabited by the study participant at index date, but also in previous homes). The results range, again, from increased lung cancer risk due to radon to no risk. Lubin et al. (1995) reminded that even these large-scale studies may not be able to provide conclusive evidence on the radon risk due to the study power being decreased by unavoidably remaining errors in exposures. However, a meta-analysis of eight case-control-studies has shown increased risk estimates (Lubin et al., 1997). To provide more conclusive evidence, the pooling of European and North-American studies is currently on-going.

It was therefore the objective of this work to clarify and specify the impact of error in exposure assessment on the radon risk for the German studies.

1.3 *The German Radon Studies*

1.3.1 Study subjects and exposure assessment

In Germany, case-control studies on lung cancer and residential radon exposure have been conducted during 1990-1998 in western and eastern parts of Germany, and primary analyses are published (Kreienbrock et al. 2001, Wichmann et al. 1998, 1999). Three study populations are covered by these studies. The first population was recruited from the areas prone to high radon concentrations in Western Germany, that is the Eifel, the Hunsrueck, Westerwald, Upper Palatinate, and Lower Bavaria (West-High). These areas are rather rural with low population density and the study was therefore extended to include bordering urban areas, now comprising the second study area (West), in order to increase the sample size. The first study population is thus a subgroup of the second. The third study area covers large parts of Saxony and Thuringia in Eastern Germany (East).

In all study areas, cases were recruited from hospitals; population controls were frequency-matched to cases on gender, age classes and geographical study regions. Each individual was personally interviewed by trained staff using a standardised questionnaire on residential, smoking and occupational history. Air measurements of radon concentrations were conducted by alpha track detectors over one year in the living room and the bedroom of the participant's homes and adjusted for house alterations and different ventilation habits of the current owner of previous homes of the participant (Gerken et al. 2000). Radon concentrations per home were computed as the average of bedroom and living room measurements weighted by individually assessed occupancy-time of rooms. In the primary analysis, radon exposure was defined as the radon concentration in the current home of the participant, and, secondly, as the participant's cumulative exposure during 5-15 years before index date. Since neither the exposure distributions nor the resulting risk estimates have been fundamentally different and the residency time of the current home has been quite long, with 23 years on average, the focus here is on the first definition (radon concentration in the current home).

1.3.2 Descriptive statistics and naive odds ratio estimates

Our analysis is based on 1449 cases and 2297 controls from the West study, including 365 cases and 595 controls from West High, and 1053 cases and 1667 controls from the East study as in the publications of the results from the primary analysis. Note that the analysed sample sizes for West and East are quite large, even though very strict inclusion criteria were adopted in order to minimise errors in exposures: The included subjects had no occupational exposure to radon, they have spent at least 25% of their time in their homes, radon measurements are available (to avoid error from imputation) and conducted for 12 ± 2 months (to avoid seasonal effects), and questionnaire information regarding smoking and residential history is complete.

Descriptive analysis and odds ratio (OR) estimates, which are approximate estimates of the relative risk (RR) of disease (Breslow, 1985), are summarised in *Table 1*. OR estimates per 100 Bq/m³ based on observed continuous radon exposure and asymptotic 95% confidence intervals are computed by logistic regression conditional on matching criteria (Using PROC PHREG by SAS®.). Controlling for smoking and occupational exposure was obtained by adjusting for the continuous variable packyears+1 and for five binary variables coding for categories of years since quitting smoking, for smoking of other products, and for occupational asbestos exposure.

Note that the exposure mean and variance are rather small, particularly for the West study, and that the OR estimates in the table deviate slightly from the published estimates, which is due to the fact that, in the published analysis, controlling for cigarette smoking was achieved by adjusting for log(packyears+1) instead of packyears+1 on the original scale (The reason for this single deviation from the primary analysis will become clear in the following.).

The correlation coefficients and the p-values of the test for equality to zero show that there is basically no correlation between radon and smoking in West-High, whereas there is a small but clear negative correlation in the West and the East study. Note that the Pearson correlation coefficient for the variables on the original scale is of no use, due to the non-normality of the distribution of radon exposure and packyears, and is only reported for completeness.

Table 1: German radon studies: Descriptive statistics of observed exposures and „naive" odds ratio estimates per 100 Bq/m³ (OR) and 95% confidence intervals (CI)

	German West *controls / cases*	*German West high* *controls / cases*	*German East* *controls / cases*
number of subjects	2297 / 1449	595 / 364	1667 / 1053
radon exposure (Bq/m³):			
mean	49 / 47	59 / 65	74 / 74
standard deviation	47 / 39	66 / 48	107 / 116
coefficient of variation	0.95 / 0.83	1.11 / 0.74	1.45 / 1.57
geometric mean	40 / 39	47 / 54	56 / 55
geometric std. dev.	1.81 / 1.78	1.84 / 1.80	1.89 / 1.98
packyears[1]:			
number of smokers[2]	1538 / 1345	390 / 339	1095 / 948
mean	17 / 35	15 / 33	12 / 25
standard deviation	20 / 21	18 / 20	14 / 16
coefficient of variation	1.19 / 0.60	1.19 / 0.59	1.11 / 0.63
geometric mean	6 / 25	6 / 24	5 / 17
geometric std. dev.	4.71 / 3.03	4.57 / 2.96	4.16 / 3.24
corr.coeff.[3] (p-value):			
Pearson, log-scale	-0.06(<0.01) / -0.01(0.59)	0.00 (0.98) / -0.01 (0.83)	-0.06(<0.01) / -0.08(0.01)
Spearman	-0.07(<0.01)/-0.05(0.04)	-0.01(0.82) / -0.02 (0.63)	-0.06 (<0.01) /-0.07(0.02)
Pearson, orig. scale	-0.02 (0.34)/-0.04 (0.14)	0.05(0.20) / 0.00 (0.98)	-0.02(0.26) / -0.07 (0.02)
OR (CI)[4]			
adjusted[5]	0.97 (0.81, 1.15)	1.12 (0.87, 1.43)	1.04 (0.96, 1.12)
raw	0.88 (0.74, 1.04)	1.15 (0.92, 1.44)	1.00 (0.93, 1.08)

[1] All statistics are given for packyear s+ 1.

[2] Current and ex-smokers of cigarettes.

[3] Between radon exposure and packyears+1 on the log-scale or on the original scale.

[4] Derived by logistic regression stratified for the matching variables.

[5] Adjusted for the continuous variable packyears+1 and five binary variables coding for categories of years since quitting smoking, smoking of other products, and occupational exposure to asbestos.

Figure 1 shows the univariate frequency distribution of (a,b) observed radon exposure and (c,d) observed packyears for controls and cases, respectively, of the West study (The distributions for West High and East are very much alike.). It can be seen that the distributions are right-skewed and reasonably approximated by the lognormal distribution.

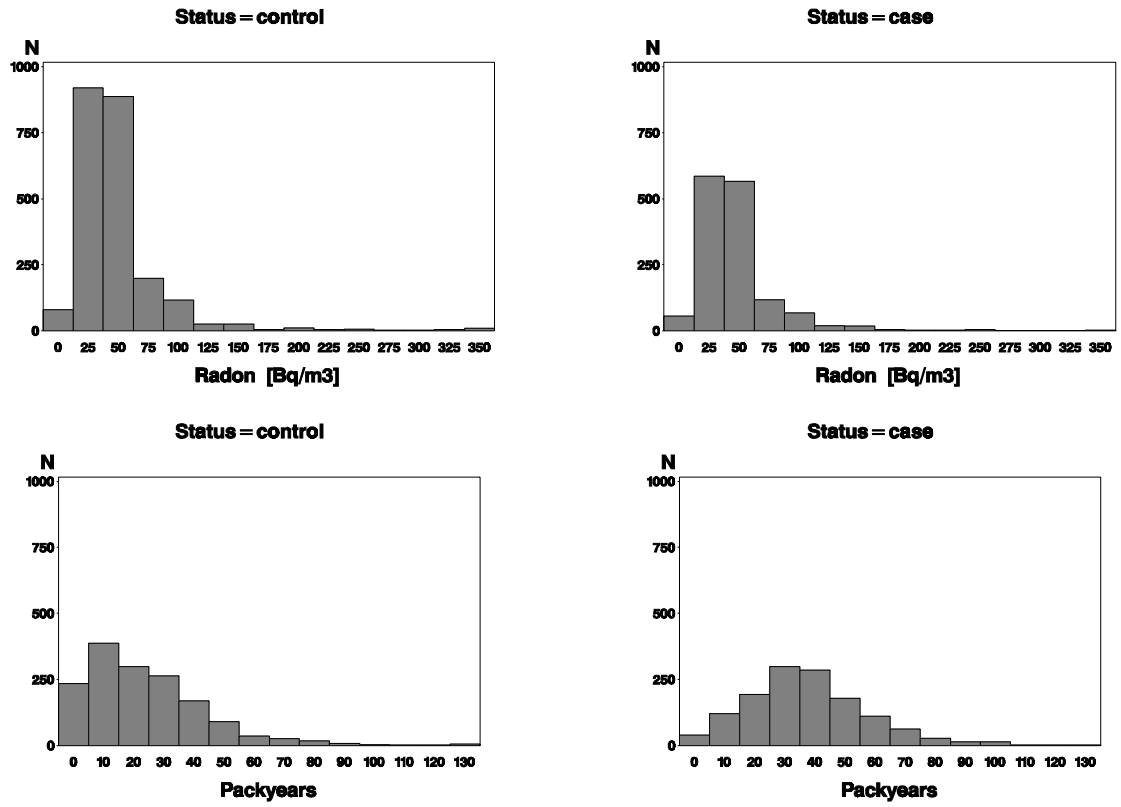

Figure 1: German West radon study: Univariate frequency distribution for (a,b) radon exposure and (c,d) packyears observed among controls and cases, respectively.

1.3.3 Contrasting the German West study region with West-High

The difference of the OR estimates (see *Table 1*) reported for the entire study region of the West study (West) and the subgroup from radon-prone matching areas (West-High) asks for explanation. The definition of the West study was initially based on administrative districts, using information of a radon survey conducted in houses in the 1980s (BMU 1992) and the underlying geology (Kemski et al., 1996). This definition yielded the radon prone areas of Eifel, Westerwald, Hunsrueck, the Upper Palatinate, and Lower Bavaria (Kreienbrock et al., 2001), which are mainly rural with low population density. To ensure the necessary study size, the more densely populated surrounding urban/metropolitan areas (among them Cologne with > 1.000.000, Düsseldorf, Essen and Dortmund with > 500.000 inhabitants) have been included subsequently. The urban/metropolitan areas exhibit particularities: Although there

are hilly/mountainous areas with occasionally higher radon, the vast majority of the population lived in the valleys with low radon. Further, these areas are highly industrialised with a more heavily smoking population. This makes the estimates of lung cancer risk due to radon exposure subject to confounding, if smoking is not properly controlled for. There is also a potential of occupational exposure to asbestos in the urban areas. However, occupational asbestos exposure has shown no notable impact on the estimates of lung cancer risk due to radon as the confounder smoking.

The uncertainty in the exposure assessment has been given as a possible explanation for the discrepant results of West and West-High. Kreienbrock et al. (2001) have reasoned that adding a large number of low-exposed subjects could possibly have decreased the exposure variance and thus increased the impact of such errors in exposures to an extent that an existent radon risk could have been masked by these effects in the entire region of the West study, whereby the true underlying risk might show in the more homogeneous region of West-High. It has been therefore been of great interest and our objective to compare the impact of measurement error in exposure assessment on the risk estimates of the West and the West-High study region.

1.3.4 Contrasting the German West and the East study

The West and the East study are part of a currently on-going pooling of European radon studies, namely studies from Sweden, Finland, United Kingdom, France, the Chech Republic, Italy, Austria and Luxembourg. The impact of measurement error is one of the issues that will be addressed in the analysis of the pooled data. It is therefore our objective to compare the impact of errors in exposure on the risk estimates of the two German studies.

1.4 Accounting for measurement error in exposures

Statistical methodology shows that random errors in exposure lead to biased relative risk (RR) estimates (Rosner et al. 1989). If the exposure is measured with classical measurement error, the RR estimates are usually underestimated. Further, such errors induce differential bias

when studies with different exposure distribution parameters are compared, even if the error and the underlying true risk is assumed to be the same (Heid et al., in press). This is, in fact, the case, when West and West-High are contrasted or if the results of several studies on the same research topic are compared. Finally, the importance of controlling for the confounder smoking was already emphasised. The question arises, how errors in this assessed confounder impact the RR estimate of lung cancer due to radon. It has been stated that residual confounding due to errors in the confounder smoking were a major threat in epidemiological studies (Michels, 2001). However, the precise effect of such errors has not yet been evaluated for epidemiological studies.

Methods to correct for such errors are abundant (see review of Thürigen et al., 2001). *Regression calibration* is one of these methods and has been investigated in great detail (e.g. Rosner et al. 1989, Carroll et al. 1994). Though this method was developed for multiple logistic regression accounting for errors in more than one predictor, application is usually restricted to account for errors in the primary predictor of interest (Darby et al. 1998, Lagarde et al 1997). Of particular interest is thus the the influence of the correlation between radon exposure and smoking, which determines the influence of errors in the confounder on the radon risk.

The error model, that is a mathematical formulation of the deviation of the measured exposure from the true exposure, is the most important model assumption, upon which all of the correction methods rely. In fact, assumptions on the error model are a particular source of concern regarding measurement error correction (Michels, 2001). The additive error model is widely discussed and applied (e.g. Armstrong et al. 1990, Ibibarren et al. 1996). However, there is less work on the multiplicative error model, which is appropriate for many exposures encountered in epidemiology with radon exposure being amongst them. The correction formulas for multiplicative error are more complex than those for additive error and are evaluated analytically to enable the differentiation between correction effects deducably from theory and effects inherent in the data.

Not only the differentiation between additive and multiplicative error models, but also the classification of the error as classical type (error independent from true exposure) or Berkson type (error independent from measured exposure) is crucial for the error correction methods.

For this reason, the single sources of error in the risk factors particularly applicable for the German radon studies were identified and classified.

For two radon studies from Sweden and the United Kingdom, corrections for multiplicative classical type error from uncertainties in the measurement of radon gas concentration in the home and extrapolation to long-term radon concentrations have been undertaken (Lagarde et al. 1997, Darby et al. 1998). Lagarde et al. (1997) have applied the regression calibration method. Darby et al. have applied an approach described by Reeves et al. (1998), additionally accounting for a Berkson type error from imputing missing measurements by the local area mean.

In this work, the error model applied by Darby et al. (1998) and Lagarde et al. (1997) is extended by not only accounting for errors in the radon exposure measurements, but also for errors in the assessment of the most potent confounder, smoking. The results obtained by the regression calibration method are contrasted with results obtained by the method described by Reeves et al. (1998).

1.5 Objectives

In the primary analysis of German radon studies involving together over 2500 cases and almost 4000 controls, OR estimates for 100 Bq/m³ increase in radon exposure were derived by logistic regression adjusted for smoking without correcting for unavoidable imprecision in exposure. It is our objective to investigate the impact of measurement error in exposure on the OR estimates.

For this work, it is thus our main objective to answer the following the questions:

1. What is a concise and practical model for the error in assessing residential radon exposure and smoking?
2. What is the impact of that error on the estimate of relative lung cancer risk from radon?
3. What is the impact of error in the confounder, smoking, on the estimate of relative lung cancer risk from radon?

4. Can the error model be extended to account for the error made by using environmental exposure as a surrogate for effective lung dose?

Further objectives include the following:

5. How does the choice of error model affect the correction of relative risk estimates?
6. Why can measurement error in exposure lead to differential bias between groups?
7. Can the observed discrepancy between the results of the full analysis and a subgroup analysis in the German West study be explained by measurement error in exposures?

In order to provide confidence into the applied correction method, it is a further goal

8. to make the method to correct for errors in exposures transparent and interpretable,
9. to particularly elucidate the role of the correlation between radon exposure and smoking.

2 MATERIAL AND METHODS

2.1 Measurement error correction methods

Methods to correct for measurement error deal with the imprecision in the assessment of the predictor variable and attempt to handle the resulting impact on the RR estimates of disease. These methods therefore touch one of the most important concepts in epidemiology: validity and precision of the obtained estimates.

The *validity* of an estimate (for example an estimate for mean radon concentration in the population or an estimate of the relative lung risk due to some risk factor) refers to the agreement between this estimate and the true underlying quantity. A statistical quantification for the violation of validity is the *bias* of the estimate, that is the true quantity and the deviation between the average estimate that would be found if many such studies were conducted in the same fashion. The *precision* of such an estimate refers to its repeatability. Errors from sampling and analysis, but also the true variability of the quantity under study in time and space worsen the precision of an estimate.

In the risk estimation in epidemiological studies, the investigator attempts to quantify the impact of a risk factor (also named „exposure“ or „predictor“) on the risk of disease. Thus, there are two notions of validity and precision to be considered in such studies. Firstly, there is the validity of the measure of the predictor, which is affected by systematic errors in the measurements (inaccurate measurements) that is an overall under- or overestimation of the predictor. The precision of the measure of the predictor is affected by random errors in the measurements (imprecise measurements), that are errors which are zero on average. Secondly, there is the validity of the risk estimate. One of the reasons for violated validity of the risk estimate is *information bias*, that is the bias in the risk estimate from inaccurately or imprecisely measured exposure. Another source of violating the validity is the *confounding bias*, that is the confounding of the effect of primary interest by the effect of another risk factor. If the confounding variable is adjusted for, but the confounder is imprecisely measured, the risk estimate of primary interest might still be biased by residual confounding. The precision of the risk estimate is usually quantified by the standard error of the risk estimate or by the width of its confidence intervals. This precision is affected by random

errors in sampling and analysis, but also by random errors in the measured exposure.

Methods to correct for errors in exposures thus have two objectives: To allow for predictors measured with random error in the risk estimation and thus provide unbiased risk estimates (pertain validity of risk estimate) and valid confidence intervals (give correct measure of the precision of the risk estimate).

2.1.1 The tool box

Correction for measurement error in exposure generally requires three models: the *model for true exposure*, the *true exposure-disease model* linking true exposure and disease, and the *error model* linking true exposure and measured exposure. Given these models, the *exposure-disease model accounting for errors in exposure* can be derived linking measured exposure and disease.

2.1.1.1 Model for the distribution of the true predictors

Let X_1 and X_2 represent two lognormally distributed true predictors of lung cancer, $X_i \sim LN(\mu_{\log X_i}, \sigma^2_{\log X_i})$, such as truly experienced radon exposure and truly smoked packyears, in one matching stratum. Thus, the logarithms of X_1 and X_2 are normally distributed with expectation $\mu_{\log X_i}$ and variance $\sigma^2_{\log X_i}$; estimates of the lognormal distribution parameters are given by the sample mean and sample variance of $\log X_i$, i=1,2.

The parameters of location and spread of the lognormal distribution on the original scale are the geometric mean (GM) and the geometric standard deviation (GSD), $\exp(\mu_{\log X_i})$ and $\exp(\sigma_{\log X_i})$. The GM is the median and the mode of the lognormal distribution. Expectation and standard deviation (SD) on the original scale are given by

$$E[X_i] = \exp(\mu_{\log X_i} + 0.5\sigma^2_{\log X_i}) \text{ and } SD[X_i] = E[X_i] \cdot \sqrt{\exp(\sigma^2_{\log X_i} - 1)},$$

i=1,2. . The coefficient of variation, CV, that is the SD divided by the expectation, is thus

$\sqrt{\exp(\sigma^2_{\log X_i} - 1)}$ and depends only on the GSD. The expectation, which is the centre of gravity of the area under the lognormal density function, and the SD increase with increasing GM or GSD. *Figure 2* shows the lognormal density function for various values of GM and GSD.

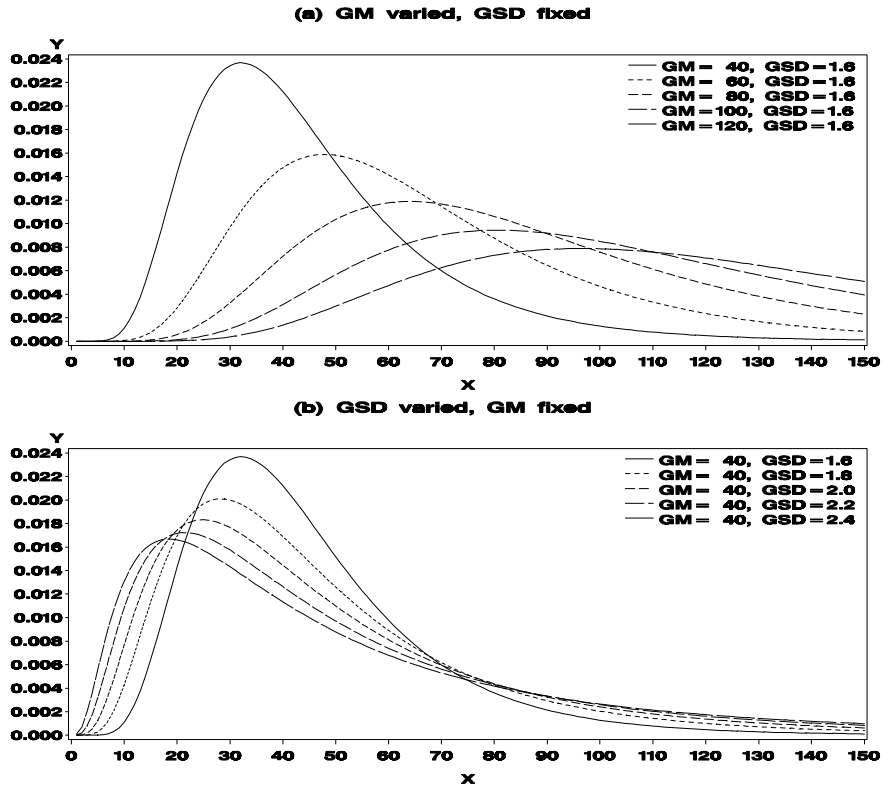

Figure 2: Model for the true predictor X: The density function of lognormal distributions with (a) varying geometric mean (GM) and (b) varying geometric standard deviation (GSD).

Vice versa, rewriting GM and GSD as

$$\mathrm{GM}[X_i] = \frac{\mathrm{E}[X_i]}{\sqrt{\frac{\mathrm{SD}^2[X_i]}{\mathrm{E}^2[X_i]} + 1}} \quad \text{and} \quad \mathrm{GSD}[X_i] = \sqrt{\log\left(\sqrt{\frac{\mathrm{SD}^2[X_i]}{\mathrm{E}^2[X_i]} + 1}\right)},$$

it can be seen that the GM increases with increasing expectation and decreasing SD, whereas the GSD increases with increasing SD.

It is further noteworthy that the product of two lognormally distributed variables is lognormal (as the sum of two normally distributed random variables is normal).

2.1.1.2 The true exposure-disease model

The logistic model involving two predictors is given by

$$\boxed{\text{Logit}(p) = \alpha + \beta_{X_1} X_1 + \beta_{X_2} X_2 + \varepsilon} \qquad (1)$$

with p being the probability of disease given the exposures X_1 and X_2, $\text{Logit}(p) = p/(1-p)$, β_{X_1} and β_{X_2} being the effect parameters corresponding to X_1 and X_2, α being a nuisance parameter in case-control studies, and ε the regression residual (Breslow and Day 1980). The OR per 100 Bq/m³ increase in radon exposure, $OR_{X_1} = \exp(100 \cdot \beta_{X_1})$, is of primary interest. If the true predictors were known without error for all individuals, the true effect parameters β_{X_1} and β_{X_2} could directly be estimated using (1).

In practice, the observed predictors, Z_1 and Z_2, such as measured radon exposure and reported packyears, are assessed with a certain imprecision. In the analysis of epidemiological studies, these observed predictors are usually plugged into the true exposure-disease model (1) without modification,

$$\boxed{\text{Logit}(p) = \alpha + \beta_{Z_1} Z_1 + \beta_{Z_2} Z_2 + \varepsilon}. \qquad (2)$$

In terms of measurement error methodology, this is a "naive" exposure-disease model and the estimates of β_{Z_1} and β_{Z_2} derived from this model are possibly biased estimates for β_{X_1} and β_{X_2}. *Figure 3* illustrates the connection between the true and naive exposure-disease model.

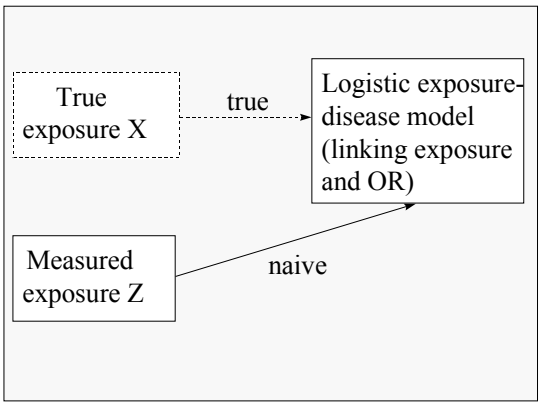

Figure 3: True and naive exposure-disease model.

2.1.1.3 Error model

In order to describe the deviation of the observed predictor from the true predictor, it is helpful to consider five classifying characteristics of error models: (1) random vs. systematic error, (2) homoscedastic vs. heteroscedastic error, (3) non-differential vs. differential error, (4) classical vs. Berkson error, (5) additive vs. multiplicative error.

A *random* error is an error with zero expectation, that is the difference between measurement and truth is on average zero. A *systematic* error has non-zero expectation. A certain random error is usually unavoidable; systematic error can be avoided by calibrating the measuring procedure. A *homoscedastic* error is an error with the same structure and from one distribution for all individuals. A *heteroscedastic* error is an error with structure or distribution differing among individuals. If the measuring procedure is the same for all individuals and there is no additional information used for some of them (e.g. the knowledge that the measurements have been conducted in different laboratories with varying calibration procedures), the error is usually homoscedastic. A *non-differential* error towards disease status has the same distribution for cases and controls. The distribution of a *differential* error differs between cases and controls. That the error in the predictor is non-differential is not trivial particularly in case-control studies, since the disease status is known throughout the exposure assessment, and blinding towards case-control status may not always be feasible.

The *classical* error is statistically independent from the true predictor variable. A classical

model is reasonable for the difference between the measured levels of an environmental agent (measured by a technical device of a certain imprecision) and the true level of that agent. The measured values spread about the true value, if the measurement is repeated. The classical error is the error that comes to mind intuitively, when thinking about „measurement error". Less intuitive at first glance is the *Berkson* error, an error which is independent of the observed predictor. A Berkson model is applicable, when one value (observed predictor value) is assigned to a group of persons, whose individual true predictor values vary. If such an observation is repeated (under the same conditions and assuming the absence of classical type error), the observation does not change. A Berkson error typically occurs, for example, by using the entry of a job-exposure-matrix, which is usually the mean exposure of a defined group of employees that are monitored and which is then applied to other employees of the same job category. Another example is the use of a monitoring device, which is on a fixed location, to assess the level of an environmental factor that is then applied for all persons living in a certain radius of that monitor.

Classical and Berkson models represent two extremes of a full range of models. Most errors in the predictor are a combination of both: In the example of the job-exposure-matrix, the matrix entries are subject to classical error since the single exposures, whose mean become the entries, are measured only with a certain error. In the example of the fixed monitors, the monitoring device cannot measure the environmental factor but with a certain imprecision, again a classical error.

The *additive* error model is a model, where the error operates additively on the exposure. The *multiplicative* error operates multiplicatively. Mostly, the additive error is assumed to be normally distributed, the multiplicative error to be lognormally distributed. The SD of the additive error is constant across the full range of the predictor. The SD of the multiplicative error is close to zero for small levels of the predictor and increases proportionally with the predictor level (Note that a lognormally distributed error is undefined for zero exposure, since the log of zero is minus infinity.).

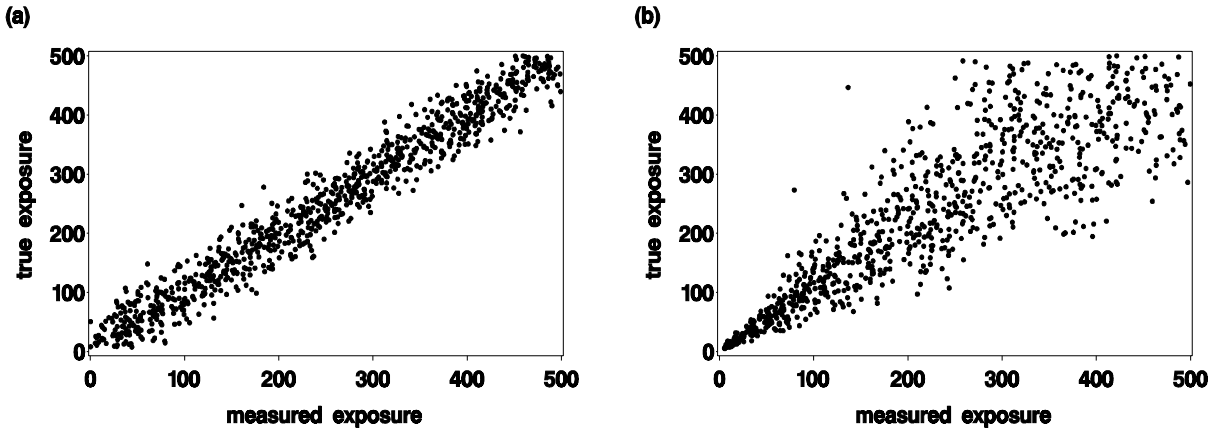

Figure 4: Error models: True predictor versus predictor measured with (a) additive or (b) multiplicative error (Fictitious data). .

The relationship between true exposure and exposure measured with additive or multiplicative classical type error (fictitious data) is displayed in *Figure 4* (For Berkson type error, the role of true and measured exposure just need to be exchanged.). It can be seen that the spread of true exposure given measured exposure (vertical spread of dots) is constant for the full range of measured exposure for the additive error model: The graph shows a „tube". For the multiplicative error model, the spread increases proportionally to measured exposure: Graph shows a „trumpet".

Since the multiplicative error is additive on the log-scale, all characteristics of the additive error are valid for the multiplicative error on the log-scale: For example, the plot of the log of true exposure versus the log of exposure measured with multiplicative error would provide the same picture as *Figure 4a*; a random multiplicative error has zero expectation on the log-scale (The expectation is not zero on the original scale!). However, as the exposures enter the exposure-disease model on the original scale, differentiation of these models is necessary.

The mathematical formulation of the additive/multiplicative error of the classical/Berkson type is summarised in *Figure 5*.

	additive	*multiplikative*
classical	**Z = X + EA** X, EA independent EA ~ N(0,σ_{EA})	**Z = X· EM** X, EM independent EM ~ LN(0,σ_{logEM})
Berkson	**X = Z + EAB** Z, EAB independent EAB ~ N(0,σ_{EAB})	**X = Z · EMB** Z, EMB independent EMB ~ LN(0,σ_{logEMB})

Figure 5: Formulation of additive/multiplicative classical/Berkson type error with X denoting the true predictor and Z the observed predictor.

In this work, a random homoscedastic error, which is non-differential towards disease status, is considered. A multiplicative error model combining a classical and Berkson component for the two predictors of lung cancer, radon and smoking, is adopted. Such a bivariate multiplicative error model can be written as

$$Z_1 = Q_1 \cdot EM_1, \qquad Z_2 = Q_2 \cdot EM_2 \qquad (3)$$
$$X_1 = Q_1 \cdot EMB_1, \qquad X_2 = Q_2 \cdot EMB_2$$

Hereby, EM_1 and EMB_1 represent the multiplicative error of the classical and the Berkson type for the predictor of primary interest, radon, with $logEM_1$ and $logEMB_1$ having expectation zero and variances $\sigma^2_{\log EM_1}$ and $\sigma^2_{\log EMB_1}$; Q_1 and Q_2 are interim variables. EM_2 and EMB_2 represent the analogous error for the confounding variable, smoking, with $\sigma^2_{\log EM_2}$ and $\sigma^2_{\log EMB_2}$ being defined analogously. The error components, EM_1, EMB_1, EM_2 and EMB_2, are assumed as statistically pair-wise independent, and the true predictors, X_1 and X_2, as pair-wise independent of the error components. Note that it is not assumed that X_1 and X_2 are independent. This error model allows to handle pure classical type error, pure Berkson type error or a combination of both.

It was already stated that lognormality of true exposures is assumed (see 2.1.1.1) and that the

data suggests that the distributions of observed predictors are reasonably well described by the lognormal distribution, LN($\mu_{\log Z_i}, \sigma^2_{\log Z_i}$), for i=1,2 (see *Figure 1*). This is in accordance with lognormally distributed errors, LN($0, \sigma^2_{\log EM_i}$) and LN($0, \sigma^2_{\log EMB_i}$). The distribution parameters of observed exposure, $\mu_{\log Z_i}$ and $\sigma^2_{\log Z_i}$, are estimated in case-control studies by the sample mean and variance of the log of observed exposure among controls. The error sizes, $\sigma^2_{\log EM_i}$ and $\sigma^2_{\log EMB_i}$, can either be assumed as exactly known, assumed as known and varied over a plausible range (sensitivity analysis), or estimated by validation, replication or instrumental data (Carroll, 1995).

For comparison, a bivariate additive error model,

$$Z_1 = Q_1 + EA_1, \qquad Z_2 = Q_2 + EA_2 \qquad (3^*)$$
$$X_1 = Q_1 + EAB_1, \qquad X_2 = Q_2 + EAB_2.$$

is adopted additionally. Hereby, EA_i and EAB_i have expectation zero and variances $\sigma^2_{EA_i}$ and $\sigma^2_{EAB_i}$, for i=1,2, and represent the additive error of classical and Berkson type, respectively, for the two variables. The error components, EA_1, EAB_1, EA_2 and EAB_2, are assumed as statistically pair-wise independent, and true predictors, X_1 and X_2, as pair-wise independent of the error components. The distribution parameters of observed predictors, μ_{Z_i} and $\sigma^2_{Z_i}$, are estimated in case-control studies by the sample mean and variance of the observed exposure on the original scale among controls. The error sizes, $\sigma^2_{EA_i}$ and $\sigma^2_{EAB_i}$, can, again, be assumed as exactly known, assumed as known and varied over a plausible range (sensitivity analysis), or estimated by validation, replication or instrumental data.

2.1.2 Regressing the true predictors on the observed predictors

The reliability coefficients are given as the regression coefficients from regressing true exposures on measured exposures.

2.1.2.1 Ignoring the association between predictors

Let us assume for now, that the true predictors, X_1 and X_2, are independent. Since the error in one predictor is independent of the error in the other, also X_1 and Z_2 [and X_2 and Z_1] are independent, that is there is no information about X_1 in Z_2 [about X_2 in Z_1].

For the multiplicative error model (3), the regression of $\log X_i$ on both $\log Z_1$ and $\log Z_2$ is thus the same as the regression on $\log Z_i$, for i=1,2. The corresponding regression coefficients are

$$\gamma_i = 1 - \sigma^2_{\log EM_i} / \sigma^2_{\log Z_i}, \quad (4)$$

and

$$\sigma_{ER_i} = \sigma^2_{\log EMB_i} + \sigma^2_{\log EM_i}\gamma_i \quad (5)$$

are the variances of the regression residuals.

For the <u>additive error model</u> (3*), the regression of X_i on Z_1 and Z_2 is the same as the regression on Z_i, for i=1,2. The corresponding regression coefficients are

$$\gamma_i = 1 - \sigma^2_{EA_i} / \sigma^2_{Z_i}, \quad (4^*)$$

and

$$\sigma_{ER_i} = \sigma^2_{EAB_i} + \sigma^2_{EA_i}\gamma_i \quad (5^*)$$

are the variances of the regression residuals.

2.1.2.2 Accounting for the association between predictors

It is now allowed for the true predictors being correlated. The error components have been assumed as independent with each other and each exposure as independent with the error of the other (see Section 2.1.1.3). This implies that the covariance between predictors (for multiplicative error between the log of exposures) is invariant to random errors.

Thus, given a <u>multiplicative error model</u> (3), the covariance between the log of true predictors, $\mathrm{cov}(\log X_1, \log X_2)$, is identical to the covariance between the log of observed predictors, $\mathrm{cov}(\log Z_1, \log Z_2)$. It can thus be directly estimated from the observed data and shall be denoted by r. Note that an estimate for the correlation between the log of observed exposures,

$\rho_{\log Z_1,\log Z_2}$, is also given by the data and that the correlation corrected for the error in predictors („true“ correlation) is given by

$$\rho_{\log X_1,\log X_2} = \frac{\rho_{\log Z_1,\log Z_2}}{\sqrt{\gamma_1 \cdot \gamma_2}} \qquad (6)$$

with γ_1 and γ_2 from (4). Given the observed correlation, the absolute value of the true correlation thus increases with increasing error size, since the γ_i's are positive and smaller than unity (The observed correlation coefficient is deflated by the error.). The regression coefficients from regressing $\log X_i$ on $\log Z_j$ (i,j=1,2) are

$$\gamma_{11} = \frac{(\sigma^2_{\log Z_1} - \sigma^2_{\log EM_1})\sigma^2_{\log Z_2} - r^2}{\sigma^2_{\log Z_1}\sigma^2_{\log Z_2} - r^2}, \quad \gamma_{12} = \frac{r\sigma^2_{\log EM_1}}{\sigma^2_{\log Z_1}\sigma^2_{\log Z_2} - r^2}, \qquad (7)$$

$$\gamma_{21} = \frac{r\sigma^2_{\log EM_2}}{\sigma^2_{\log Z_1}\sigma^2_{\log Z_2} - r^2}, \qquad \gamma_{22} = \frac{(\sigma^2_{\log Z_2} - \sigma^2_{\log EM_2})\sigma^2_{\log Z_1} - r^2}{\sigma^2_{\log Z_1}\sigma^2_{\log Z_2} - r^2},$$

and

$$\sigma_{ER_{11}} = \sigma^2_{\log EMB_1} + \sigma^2_{\log EM_1}\gamma_{11}, \quad \sigma_{ER_{12}} = \sigma^2_{\log EM_2}\gamma_{12}, \qquad (8)$$

$$\sigma_{ER_{21}} = \sigma^2_{ER_{12}}, \qquad \sigma_{ER_{22}} = \sigma^2_{\log EMB_2} + \sigma^2_{\log EM_2}\gamma_{22}$$

denote the elements of the covariance matrix of the regression residuals.

For the <u>additive error model</u> (3*), the same is valid on the original scale instead of on log-scale: The covariance between true predictors is identical to the covariance between observed predictors. It is denoted by r and estimated from the observed data. The „true“ correlation is

$$\rho_{X_1,X_2} = \frac{\rho_{Z_1,Z_2}}{\sqrt{\gamma_1 \cdot \gamma_2}} \qquad (6^*)$$

with γ_1 and γ_2 from (4*). The regression coefficients from regressing X_i on Z_j (i,j=1,2) are

$$\gamma_{11} = \frac{(\sigma^2_{Z_1} - \sigma^2_{EA_1})\sigma^2_{Z_2} - r^2}{\sigma^2_{Z_1}\sigma^2_{Z_2} - r^2}, \quad \gamma_{12} = \frac{r\sigma^2_{EA_1}}{\sigma^2_{Z_1}\sigma^2_{Z_2} - r^2}, \qquad (7^*)$$

$$\gamma_{21} = \frac{r\sigma^2_{EA_2}}{\sigma^2_{Z_1}\sigma^2_{Z_2} - r^2}, \qquad \gamma_{22} = \frac{(\sigma^2_{Z_2} - \sigma^2_{EA_2})\sigma^2_{Z_1} - r^2}{\sigma^2_{Z_1}\sigma^2_{Z_2} - r^2},$$

and

$$\sigma_{ER_{11}} = \sigma^2_{EAB_1} + \sigma^2_{EA_1}\gamma_{11}, \quad \sigma_{ER_{12}} = \sigma^2_{EA_2}\gamma_{12}, \qquad (8^*)$$

$$\sigma_{ER_{21}} = \sigma^2_{ER_{12}}, \qquad \sigma_{ER_{22}} = \sigma^2_{EAB_2} + \sigma^2_{EA_2}\gamma_{22}$$

denote the elements of the covariance matrix of the regression residuals.

2.1.3 Regression calibration method to account for errors in two predictors

An exposure-disease model accounting for errors in exposures can be obtained according to the regression calibration method (Rosner et al. 1989). Hereby, the observed predictors, Z_1 or Z_2, are replaced in the naive model (2) by the expectation of the corresponding true predictor given the observed predictors, $T_1 = E[X_1 \mid (Z_1, Z_2)]$ and $T_2 = E[X_2 \mid (Z_1, Z_2)]$.

Given the logistic model as true exposure-disease model, this correction method involves computing T_1 and T_2 and estimating the beta-coefficients of

$$\boxed{\text{Logit}(p) = \alpha + \beta_{T_1} T_1 + \beta_{T_2} T_2 + \varepsilon} \qquad (9)$$

to obtain approximately unbiased estimates for the true effect parameters, β_{X1} and β_{X2}. *Figure 6* illustrates the mechanism of regression calibration.

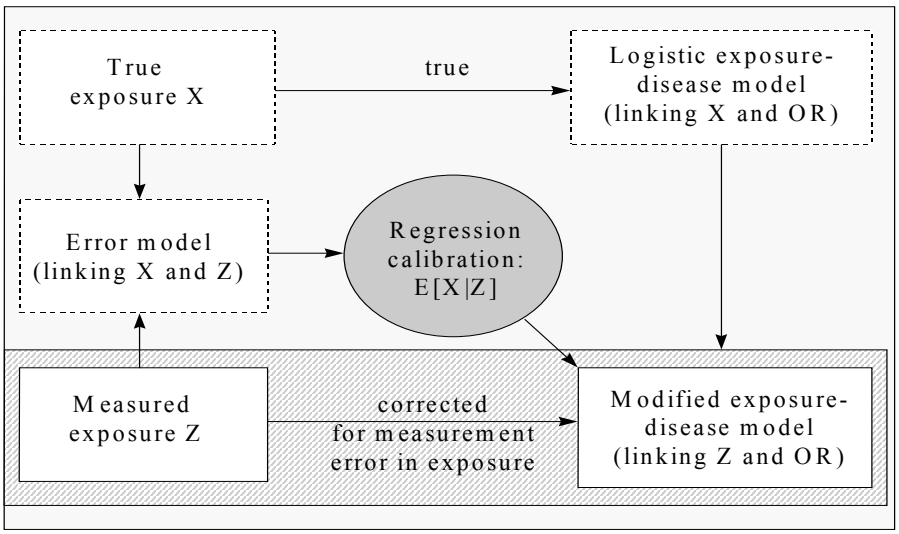

Figure 6: Regression calibration: Hatched lining indicates a theoretical model assumption, solid lining indicates what is derived from the observed data or by logical deduction, the striped rectangle summarises what is actually used in the analysis accounting for measurement error in exposure.

2.1.3.1 Ignoring the association between predictors

Let us assume for now that the true predictors, X_1 and X_2, are not associated. Since the errors are assumed to be independent (see 2.1.1.3), consequently X_1 and Z_2 as well as X_2 and Z_1 are independent. Therefore, the distribution of $X_i|(Z_1,Z_2)$ is the same as of $X_i|Z_i$ (i=1,2). The regression calibration method thus implies computing $T_1 = E[X_1|Z_1]$ and $T_2 = E[X_2|Z_2]$ to replace Z_1 and Z_2 in model (2).

The typical <u>multiplicative error model</u> implies that errors and predictors are lognormally distributed. Then, the distribution of $X_i|Z_i$ is lognormal, $LN(\mu_{\log X_i|Z_i}, \sigma^2_{\log X_i|Z_i})$ with $\mu_{\log X_i|Z_i} = (1-\gamma_i)\cdot\mu_{\log Z_i} + \gamma_i \log Z_i$ and $\sigma^2_{\log X_i|Z_i} = \sigma_{ER_i}$ (i=1,2, γ_i and σ_{ER_i} from (4) and (5)). For each observation with observed predictor, Z_i, compute T_i as $m_i \cdot Z_i^{\gamma_i}$ with $m_i = (\exp(\mu_{\log Z_i}))^{1-\gamma_i} \cdot \exp(0.5\sigma_{ER_i})$ and obtain

$$\boxed{\text{Logit}(p) = \alpha + \beta_{T_1} m_1 Z_1^{\gamma_1} + \beta_{T_2} m_2 Z_2^{\gamma_2} + \varepsilon} \qquad (10)$$

is an exposure-disease model accounting for multiplicative error in two independent predictors.

The typical <u>additive error model</u> implies that errors and predictors are normally distributed. Hence, the distribution of $X_i|Z_i$ is normal, $N(\mu_{X_i|Z_i}, \sigma^2_{X_i|Z_i})$ with $\mu_{X_i|Z_i} = (1-\gamma_i)\cdot\mu_{Z_i} + \gamma_i Z_i$ and $\sigma^2_{X_i|Z_i} = \sigma_{ER_i}$ (i=1,2, γ_i and σ_{ER_i} from (4*) and (5*)). For each observed predictor, Z_i, compute T_i as $\gamma_i Z_i$ and obtain

$$\boxed{\text{Logit}(p) = \tilde{\alpha} + \beta_{T_1}\gamma_1 Z_1 + \beta_{T_2}\gamma_2 Z_2 + \varepsilon} \qquad (11)$$

with $\tilde{\alpha} = \alpha + (1-\gamma_1)\cdot\mu_{Z_1} + (1-\gamma_2)\cdot\mu_{Z_2}$ as an exposure-disease model accounting for multiplicative error in two uncorrelated predictors.

In the case of homoscedastic additive errors, the beta coefficient estimates can be corrected without using the full data: Comparing model (11) with the naive model (2) results in

$$\hat{\beta}_{T_i} = \frac{1}{\gamma_i}\hat{\beta}_{Z_i},$$

for i=1,2. Hence, to correct the effect parameter estimates for homoscedastic additive error,

the naive model (2) can be applied and the resulting naive beta coefficient estimates can be derived by division through γ_i. The bias from such error, that is the difference between the true coefficient and the average naive estimate and can be estimated by the average difference between $\hat{\beta}_{T_1}$ and $\hat{\beta}_{Z_1}$, as given by the regression calibration method, is thus $(1-\gamma_i)\beta_{X_i}$.

2.1.3.2 Accounting for the association between predictors

It is now allowed for a non-zero correlation between the observed predictors.

For the multiplicative error model, error and predictors are typically assumed as lognormally distributed and thus the distribution of $X_i|(Z_1,Z_2)$ is also lognormal, $LN(\mu_{\log X_i|(Z_1,Z_2)}, \sigma^2_{\log X_i|(Z_1,Z_2)})$ with $\sigma^2_{X_i|(Z_1,Z_2)} = \sigma_{ER_{ii}}$ and $\mu_{\log X_i|(Z_1,Z_2)} = (1-\gamma_{ii})\mu_{\log Z_i} - \gamma_{ij}\mu_{\log Z_j} + \gamma_{ii}\log Z_i + \gamma_{ij}\log Z_j$ (i,j=1,2, $i \neq j$, γ_{ij} and $\sigma_{ER_{ii}}$ from (7) and (8)). Compute T_1 and T_2 as $m_1 Z_1^{\gamma_{11}} Z_2^{\gamma_{12}}$ and $m_2 Z_1^{\gamma_{211}} Z_2^{\gamma_{22}}$ with $m_i = (e^{\mu_{\log Z_i}})^{1-\gamma_{ii}} (e^{\mu_{\log Z_j}})^{-\gamma_{ij}} e^{0.5\sigma_{ER11}}$ (i,j=1,2 and $i \neq j$) and obtain

$$\boxed{\text{Logit}(p) = \alpha + \beta_{T_1} m_1 Z_1^{\gamma_{11}} Z_2^{\gamma_{12}} + \beta_{T_2} m_2 Z_2^{\gamma_{22}} Z_1^{\gamma_{21}} + \varepsilon} \qquad (12)$$

as exposure-disease model accounting for multiplicative error in two correlated predictors.

For the additive error model, errors and predictors are typically assumed as normally distributed, and thus the distribution of $X_i|(Z_1,Z_2)$ is $N(\mu_{X_i|(Z_1,Z_2)}, \sigma^2_{X_i|(Z_1,Z_2)})$ with $\mu_{X_i|(Z_1,Z_2)} = (1-\gamma_{ii})\mu_{Z_i} - \gamma_{ij}\mu_{Z_j} + \gamma_{ii}Z_i + \gamma_{ij}Z_j$ and $\sigma^2_{X_i|(Z_1,Z_2)} = \sigma_{ER_{ii}}$ (i,j=1,2, $i \neq j$, γ_{ij} and $\sigma_{ER_{ii}}$ from (7*) and (8*)). Thus, compute T_1 and T_2 as $\gamma_{11}Z_1 + \gamma_{12}Z_2$ and $\gamma_{22}Z_2 + \gamma_{21}Z_1$ and obtain

$$\boxed{\text{Logit}(p) = \tilde{\alpha} + \beta_{T_1}(\gamma_{11}Z_1 + \gamma_{12}Z_2) + \beta_{T_2}(\gamma_{22}Z_2 + \gamma_{21}Z_1) + \varepsilon} \qquad (13)$$

as exposure-disease model accounting for additive error in two correlated predictors.

In the case of homoscedastic additive errors, the beta coefficients can be corrected without using the full data: Rewriting model (13) as

$\text{Logit}(p) = \tilde{\alpha} + (\beta_{T_1}\gamma_{11} + \beta_{T_2}\gamma_{21})Z_1 + (\beta_{T_2}\gamma_{22} + \beta_{T_1}\gamma_{12})Z_2$ and comparing it with the naive model (2), provides that the naive effect parameter estimates, $\hat{\beta}_{Z_1}$ and $\hat{\beta}_{Z_2}$, are given by $\hat{\beta}_{T_1}\gamma_{11} + \hat{\beta}_{T_2}\gamma_{21}$ and $\hat{\beta}_{T_2}\gamma_{22} + \hat{\beta}_{T_1}\gamma_{12}$, which implies that the corrected estimates are given by

$$\hat{\beta}_{T_1} = \frac{\gamma_{22}\hat{\beta}_{Z_1} - \gamma_{21}\hat{\beta}_{Z_2}}{\gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21}} \text{ and } \hat{\beta}_{T_2} = \frac{\gamma_{11}\hat{\beta}_{Z_2} - \gamma_{12}\hat{\beta}_{Z_1}}{\gamma_{11}\gamma_{22} - \gamma_{12}\gamma_{21}}.$$

Hence, to correct the OR estimate for homoscedastic additive error, the naive model (2) can be applied and the resulting naive beta coefficient estimates can be corrected accordingly (Note that the beta coefficient estimates cannot be corrected directly for multiplicative error; the availability of the full data is required.). Thus, the quantity

$$(1 - \gamma_{11})\beta_{X_1} - \gamma_{21}\beta_{X_2} \qquad (14)$$

describes the bias in the effect parameter of primary interest, that is the average difference between $\hat{\beta}_{T_1}$ and $\hat{\beta}_{Z_1}$.

2.1.3.3 Computation of beta-coefficients and confidence intervals

If the error sizes can be assumed as known (as it is the case in a sensitivity analysis), not only the beta-coefficient estimates, but also the estimates of the standard error of these estimates and thus the corresponding confidence intervals can be derived by computing T_i and solving the usual logistic regression model based on T_i (i=1,2).

If the error size is estimated by validation, replication or instrumental data, the additional uncertainty from this estimating process needs to be taken into account for example by the Bootstrap method (Carroll, 1995, Küchenhoff et al., in preparation).

2.1.4 Approximate Likelihood Method

By utilising the approximation of the Logit by the Probit, similar exposure-disease models accounting for multiplicative or additive errors of classical and Berkson type can be derived

(Reeves et al, 1998). These models differ from the models derived by the regression calibration method only by the division of T_i by a square root, which shall be denoted by s:

$$\boxed{\text{Logit}(p) = \alpha + \beta_{T_1}\frac{T_1}{s} + \beta_{T_2}\frac{T_2}{s} + \varepsilon}. \qquad (15)$$

In the following, the squareroots are given for the various models and an algorithm to compute the beta-coefficient estimates. If the squareroot s can reasonably be approximated by unity, the two methods, regression calibration and this approximate likelihood method provide identical models.

In the following, let k be 0.0588, a constant involved in approximating the Logit by the Probit.

2.1.4.1 Ignoring the association between predictors

For <u>multiplicative error</u> in the predictors, T_i is computed as $m_i Z_i^{\gamma_i}$ (i=1,2) as in 2.1.3.1 and

$$s = \sqrt{1 + k^2\beta_1^2(e^{\sigma_{ER_1}} - 1)T_1^2 + k^2\beta_2^2(e^{\sigma_{ER_2}} - 1)T_2^2} \qquad (16)$$

with σ_{ER_i} from (5).

For <u>additive error</u>, T_i is computed as $\gamma_i Z_i$ (i=1,2) as in 2.1.3.1 and

$$s = \sqrt{1 + k\beta_1^2\sigma_{ER_1} + k\beta_2^2\sigma_{ER_2}} \qquad (17)$$

with σ_{ER_i} from (5*).

2.1.4.2 Accounting for the association between predictors

For <u>multiplicative error</u>, T_i is computed as $m_i Z_i^{\gamma_{ii}} Z_j^{\gamma_{ij}}$ (i,j=1,2 and i≠j) as in 2.1.3.1 and

$$s = \sqrt{1 + k^2\beta_1^2(e^{\sigma_{ER_{11}}} - 1)T_1^2 + k^2\beta_2^2(e^{\sigma_{ER_{22}}} - 1)T_2^2 + 2k^2\beta_1\beta_2(e^{\sigma_{ER_{12}}} - 1)T_1T_2}. \qquad (18)$$

with $\sigma_{ER_{ij}}$ from (7).

For <u>additive error</u>, Ti is computed as $\gamma_{ii}Z_i + \gamma_{ij}Z_j$ (i,j=1,2 and i≠j) as in 2.1.3.1 and

$$s = \sqrt{1 + k\beta_1{}^2\sigma_{ER_{11}} + k\beta_2{}^2\sigma_{ER_{22}} + 2k^2\beta_1\beta_2\sigma_{ER_{12}}}\ . \qquad (19)$$

with $\sigma_{ER_{ij}}$ from (7*).

2.1.4.3 Computation of beta-coefficient estimates and confidence intervals

Since the parameters β_{T_1} and β_{T_2} appear in the squareroot, their estimates need to be computed iteratively:

1. Chose a starting value for each effect estimates $\hat{\beta}_{T_1}$ and $\hat{\beta}_{T_2}$, for example $\hat{\beta}_{T_1}(0) = 1$ and $\hat{\beta}_{T_2}(0) = 1$.
2. Compute s for these beta coefficient and denote it by $s(k)$ for k=1.
3. Estimate β_{T_1} and β_{T_2} in $\text{Logit}(p) = \tilde{\alpha} + \beta_{T_1}\frac{T_1}{s(k)} + \beta_{T_2}\frac{T_2}{s(k)} + \varepsilon$ and denote them by $\hat{\beta}_{T_1}(k)$ and $\hat{\beta}_{T_2}(k)$ with k=1.
4. Repeat step 2 and 3 for k=2,3, ..., until $|\hat{\beta}_{T_i}(k) - \hat{\beta}_{T_i}(k-1)|$ become smaller than some given precision (i=1,2).

The confidence intervals cannot be computed directly from these models, even if the error size is assumed as known. An approach deriving the standard error of the beta-coefficient estimates, and thus confidence intervals, from the second derivative of the maximum likelihood is described by Reeves et al. 1998. Since the maximum likelihood for conditional logistic regression is very tedious to compute and not apt to be realised in a standard statistical software program, the Bootstrap method can be another alternative to derive appropriate confidence intervals. The Bootstrap method is computing-time intensive, but can be performed with standard statistical software such as SAS®.

2.1.5 Sensitivity analysis

The multiplicative classical error was described as suitable to model the error in radon exposure (Lubin et al., 1995, Lagarde et al., 1997, Darby et al., 1998). Since the error size is difficult to pinpoint and the correction highly dependent upon error size (Caroll et al. 1994), a sensitivity analysis, exploring the impact of various sizes of multiplicative classical errors in the risk factors on the OR estimates in the German radon studies, is conducted. Hereby, model (10) and model (12), derived by regression calibration to perform the error correction (without and with accounting for the correlation between radon exposure and smoking), are applied. The errors in radon exposure in this analysis explain 0 to about 50% of the variance of the log of radon exposures observed among controls in the West study (the study with the smallest variance of log radon). Further, errors in the smoking variable, packyears+1, explaining 0 to about 50% of the packyear variance on the log scale among controls in the East study (smallest variance of log of smoking variable) are investigated. There is one outlier in the West study, a control, which is also part of West-High; the sensitivity analysis is thus repeated with exclusion of this outlying observation.

To evaluate the dependence of the results on the choice of the error model, the analysis is repeated with an additive classical error using (11) and (13) (without and with accounting for the correlation). One note regarding the estimation of the correlation coefficient: The Pearson correlation coefficient is very sensitive to deviations of the variables' distributions from the normal distribution and does not provide a valid estimate for the association between lognormal exposures on the original scale. Thus, the analysis using the Spearman correlation coefficient is conducted, which involves only the ranks of the observations and is therefore robust against the distributions' non-normality; it is further scale-independent (no matter if computed with the variables on the original or on the log-scale). The product of the Spearman correlation coefficient and the SD of the two risk factors is thus used as estimate for the covariance between the variables, which is methodologically not completely correct, but necessary to cope with the problem that an additive error is adopted for lognormal predictors.

Further, the error model is extended to allow for Berkson error in the predictors. Finally, it is evaluated whether the approximate likelihood method from 2.1.4 provides different corrections of the ORs in the German radon studies than the regression calibration method.

2.2 Replication data

Information on structure and size of the error in a measured variable, Z, can be obtained

- by *validation data*, in which the true variable, X, is observable directly,
- by *replication data*, in which replicates of the error-prone variable are available, $Z^{(1)}$, ..., $Z^{(n)}$, or
- by *instrumental data*, in which another variable (the instrumental variable), W, is observable in addition to the primarily measured variable, Z, contributing some information about the true variable.

These data can be given for a subsample from the study population (*internal data*) or from comparable groups outside of the study population (*external data*). For external data, the transferability of the information to the study population needs to be ascertained. Validation data is usually collected, when the assessment of the true predictor or at least a „gold standard" of the exposure assessment is possible, but too expensive to be conducted for all study participants. Instrumental data is often collected, for example in nutritional studies, where information on nutrient intake are gathered by two methods, each of them contributing to the knowledge about true nutrient intake. Replication data is most often feasible, particularly when technical devices are the measuring tools.

Three data sets with replicate radon concentration measurements are available to obtain evidence on the multiplicativity and information on the size of the error in assessing radon exposures. „Error size" is the SD of the error (on log-scale).

2.2.1 Data

2.2.1.1 Bedroom and living room measurements

For each study participant in the analysed sample of the German radon studies, one-year measurements of radon gas concentrations in the bedroom and living room of the home inhabited at index date are available. This internal data for all participants provides thus repeated measurements in the same home during the same year. It allows the estimation of the between-measurement-variability, given the differences between bedroom and living room

can be controlled for: Radon concentrations differ between the rooms for reasons other than pure between-measurement-variability due to different floor levels or the fact, that the home owner ventilates the rooms differently (Gunby et al, 1979, Gerken et al. 2000). Floor numbers of the rooms and information on ventilation was gathered by interview. Since the error model is to suit the general population, the analysis is solely based on the controls.

2.2.1.2 Year-by-year replicates

Data on radon gas concentrations measured by alpha track detectors for several consecutive time periods during October 1995 and March 2001 was collected by the Federal Office of Radiation Protection to monitor the changes of radon concentrations over time (kindly provided by R. Lehmann). These measurements have been conducted in 11 arbitrarily (that is not randomly!) selected houses including basements of laboratories and extremely highly exposed houses in Schneeberg, which is an area with a history of uranium mining. Two measurements are made under completely identical conditions in each house and time period. The single time periods cover about one to two months. For each house, the first time period starts in October 1995. The time-weighted average radon concentration of a set of consecutive time periods covering 12 months is compute and the average radon gas concentration over one year is obtained for five consecutive years. This external data with two simulaneous one-year measurements of radon gas concentrations in 11 houses for five years ($2 \cdot 11 \cdot 5 = 110$ one-year measurements) allows the estimation of the between-year-variability (corrected for the between-measurement-variability) and of the between-measurement-variability.

2.2.1.3 Data from laboratory intercomparison study

During November 1990 and Mai 1991, an intercomparison study was conducted to evaluate within- and between-lab-variability of five laboratories from different European countries (Poffign, 1992, Kreienbrock et al, 1999). This intercomparison study has been initiated with regard to the planned pooling of European radon studies. Hereby, repeated six-month measurements of radon gas concentrations were made in five houses with radon

concentrations typical of those expected in the then on-going epidemiological German radon studies. In one room of each house, five detectors of each laboratory were placed for a six-month period. From this external data, the measurements from the German laboratory of the Biophysics Department of the University of Saarland (5 detectors in each of 5 houses = 25 six-month measurements) are used in order to estimate the between-measurement variability for the laboratory, which conducted all the measurements of the German radon studies.

2.2.2 Method of analysis

2.2.2.1 Bedroom and living room measurements

To justify the multiplicative structure of the error from between-measurement-variability, the bedroom and living room measurements are plotted for all controls with both rooms on the same floor versus the mean of the two measurements on the original scale and on log-scale (see 2.1.1.3).

The size of the error from between-measurement-variability is estimated by computing the variances (VAR) between the measurements in bedroom and living room by house and the square root of their average over all houses with both rooms at the same floor level,

$$\sqrt{\underset{i}{\mathrm{MEAN}}(\mathrm{VAR}(\log Z_{i,0}, \log Z_{i,1}))}\,,$$

with $Z_{i,j}$ indicating the radon concentration measurement of the ith subject in the bedroom (j=0) or living room (j=1). Additionally, ANOVA (using PROC MIXED by SAS®) is conducted modelling a fixed effect for the room,

$$\log Z_{i,j} = \mu + \mathrm{HOME}_i + \mathrm{ROOM}_j + \varepsilon_{i,j}\,,$$

with HOME_i denoting the effect for each house and ROOM_j the effect of the bedroom versus the living room (0: living room, 1: bedroom). The estimated SD of the residuals, $\varepsilon_{i,j}$, provides an estimate for the error from between-measurement-variability controlled for the between-room-difference. Further, the effect of including a fixed effect for the floor level of the room on the error size estimate is explored. A variable FLOORB [FLOORL] is coded by

0, 1 or 2, when the bedroom [living room] is at ground floor, first floor, or any upper floor, respectively, and a difference measure FLOOR is computed by FLOORB-FLOORL. Then the model

$$\log Z_{i,j} = \mu + HOME_i + ROOM_j + FLOOR_k + \varepsilon_{i,j},$$

is applied, with $FLOOR_k$ denoting the effect of the difference between floor levels, k=-2, ...2.

2.2.2.2 Year-by-year replicates

The measurements are plotted versus their mean within each house on the original scale and on the log-scale to obtain evidence on the multiplicativity of the error from between-year-variability.

The size of the error from between-measurement-variability is estimated by

$$\sqrt{\underset{i,j}{MEAN}(\underset{k}{VAR}(\log Z_{i,j,k}))},$$

the size of the error from between-year-variability corrected for the between-measurement-variability is estimated by

$$\sqrt{\underset{i}{MEAN}(\underset{j}{VAR}(\underset{k}{MEAN}\log Z_{i,j,k}))},$$

and the size of the combined error is estimated by

$$\sqrt{\underset{i}{MEAN}(\underset{j,k}{VAR}(\log Z_{i,j,k}))},$$

where $Z_{i,j,k}$ is the kth radon concentration measurement (k=1,2) in the jth year (j=1,..., 5) and the ith house (i=1,..., 11). Additionally, the ANOVA model with a fixed effect for the year and a random effect for the year within house,

$$\log Z_{i,j,k} = \mu + HOME_i + \gamma_{i,j}HOME_i \cdot YEAR_j + YEAR_j + \varepsilon_{i,j,k}$$

is applied, with $HOME_i$ denoting the effect of the ith house (i=1, ...,11), $YEAR_j$ a fixed effect of the year independent from home, and $HOME_i \cdot YEAR_j$ the effect in the jth year by home (j=1, ..5). An estimate for the size of the error from between-year-variability is then

given by the squareroot of the estimate of $\mathrm{HOME}_i \cdot \mathrm{YEAR}_j$, and an estimate for the size of the error from between-measurement-variability by the squareroot of the residual variance estimate. Further, it is tested whether the fixed effect of the year is statistically significant. If not, the analysis is repeated without this fixed effect. Finally, the model

$$\log Z_{i,j,k} = \mu + \mathrm{HOME}_i + \varepsilon_{i,j,k}$$

is applied to derive an estimate of the size of the combined error (error from between-year-variability and from between-measurement-variability) by the squareroot of the residual variance estimate.

2.2.2.3 Data from laboratory intercomparison study

The measurements are plotted versus their mean within each house on the original scale and on the log-scale to provide evidence for the multiplicativity of the error from between-measurement-variability.

The size of the error from between-measurement-variability is estimated by

$$\sqrt{\underset{i}{\mathrm{MEAN}}\left(\underset{j}{\mathrm{VAR}}(\log Z_{i,j})\right)},$$

where $Z_{i,j}$ is the jth radon concentration measurement (j=1,.., 5) in the ith house (i=1,..., 5). The same estimate should be obtained by applying the ANOVA model

$$\log Z_{i,j} = \mu + \mathrm{HOME}_i + \varepsilon_{i,j},$$

with HOME_i denoting the effect of the ith house (i=1, ...,5), and computing the squareroot of the residual variance estimate.

3 RESULTS

In the following, the exposure-disease model accounting for errors in the risk factors is evaluated, a feasible model for the error in assessing residential radon exposure and the confounder smoking is described, and the impact of such errors on the RR estimates of the German radon studies is elucidated.

3.1 The Error Model in the German Radon Studies

In order to establish an error model, the components of error in assessing radon exposure and smoking are identified and classified with regard to their type (classical versus Berkson). Further it is determine whether a multiplicative or an additive structure is more appropriate, and it is reflected on a plausible range for the error size. The information of the replicate data is exploited to give further evidence about the error structure and size.

3.1.1 Predictor variable of interest

Before talking about errors in the assessed exposure, one needs to think about „true exposure“, or better, about what is defined as „true exposure“ What is noted as „true exposure“ in measurement error methodology is the predictor variable of true interest. At first glance, this seems trivial. However, without concrete definition, the error model and the results of the analysis may be misinterpreted.

3.1.1.1 Concentration, exposure or dose

For the statistician, „exposure“ can be any independent variable, which is plugged into the „exposure-disease-model“. The biologist uses the term „exposure“ more specifically and distinguishes between several stages of the disease-causing agent, namely concentration, exposure and dose (Armstrong et al. 1989).

The *concentration* is a measure of the density of the agent and usually dependent on time t,

$$c(t) .$$

The *exposure* accumulated during a certain time period T (*cumulative exposure*) can then be described as

$$\int_T c(t)dt ,$$

or, sometimes, the cumulative exposure is given per time unit (for example per year) in order to provide comparability across individuals with different exposure length T,

$$X(T) = \frac{\int_T c(t)dt}{T} .$$

If the time period T is identical for all individuals in a study, as it is the case in the definition of the cumulative exposure 5 to 15 years prior to index date in the German radon studies, the two definitions of cumulative exposure are equivalent.

The concentration is usually measured at some points in time, t_i, for i=1, ..., n, and is assumed to be fairly constant during the time periods T_i with $t_i \in T_i$. The integral over time is approximated by summation to provide a *proxy for the exposure*,

$$X(T)_{proxy} = \frac{\sum_i c(t_i)T_i}{\sum_i T_i} .$$

The *dose* from the exposure experienced during time period T, which is effective at the organ site, is often influenced by a row of factors, $f_1,...,f_p$,

$$D(t) = f_1 \cdot ... f_p \cdot X(t) .$$

In the following, the word „predictor“ is used instead of the statistical term of „exposure“ in order to not confuse it with the more specific biological term. The term „exposure-disease-model“ is a fixed terminology and „predictor-disease-model“ is completely unusual. The term „exposure-disease-model“ is therefore further used, even if the independent variable in the model is a „concentration“ or a „dose“.

3.1.1.2 Radon concentration, radon exposure and alpha dose

On the example of radon as lung cancer predictor, the *true radon gas concentration* is the environmental agent that is measured by alpha track detectors in the ith home inhabited by the study participant during a time period T,

$$c(t_i) = RN_i .$$

(For clarity, the extra subscript indicating the individual study participant is omitted.). The unit for radon gas concentration is *Bequerel per cubic metre* (Bq/m³). In most recent radon studies, this concentration is derived as the average between bedroom and living room concentrations, RN_i^B and RN_i^L, weighted by the relative occupancy time, that is the percentage of time spent in the bedroom (w_i^B) or the rest of the home ($1 - w_i^B$), respectively, $RN_i = 0.5 \cdot (w_i^B \cdot RN_i^B + (1 - w_i^B) \cdot RN_i^L)$.

A proxy for the residential radon exposure experienced by the study participant during the time period T per year is the *time-weighted average* (TWA) *concentration*, that is the sum of these concentrations weighted by residency time (in years),T_i, for i=1, ...n,

$$X(T)_{proxy} = TWA(T) = \frac{\sum_i RN_i T_i}{\sum_i T_i},$$

or the *cumulative radon exposure* during T per year (Rncum), which is additionally weighted by absolute occupancy time (in hours per day),O_i, for i=1, ...n,

$$X(T)_{proxy} = RNCUM(T) = \frac{\sum_i RN_i T_i O_i}{\sum_i T_i}.$$

The unit for TWA(T) or RNCUM(T) is Bequerel-years per cubic metre and year, which is equivalent to Bequerel per cubic metre (Bq a / m³ a=Bq/m³).

The quantities RN_i, T_i, and O_i cannot be observed but with a certain error. With the observed quantities denoted by RN_i^*, T_i^*, and O_i^*, the derived proxy for the radon exposure („observed“ exposure) is then given by $TWA(T)^* = (\sum_i RN_i^* T_i^*) / (\sum_i T_i^*)$ and

$RNCUM(T)^* = (\sum_i RN_i^* T_i^* O_i^*) / (\sum_i T_i^*)$, respectively.

The alpha particles of the radioactive decay in the lung determines the lung dose (*alpha dose*). The individually experienced alpha dose in the lung during time period T differs from the radon exposure by certain factors, $f_1, \ldots, f_p$,

$$D(T) = f_1 \cdot \ldots f_p \cdot X(T)$$

One of these factors is the *equivalence factor* of radon and radon progeny describing the equilibrium between radon and radon daughters given the environmental conditions (compression, temperature, etc.): The lung dose is determined rather by the radon progeny exposure than by the radon exposure, since radon has a longer half-life (3 days) than the decay products ^{218}Po (3 minutes) and ^{214}Po (164 micro seconds); therefore, radon is often disposed of before the alpha decay occurs in the lung. Another factor is given by the difference between the environmental radon (and radon progeny) exposure and the amount that is actually inhaled and enters the human body, which depends on particles in the air (dust and smoke) and the individual's respiratory characteristics (inhalation depth and frequency and bronchial geometry). A further factor takes into account that the effective alpha dose depends also on the individual's ability to quickly dispose of the alpha particles in the bronchial system.

3.1.1.3 Number of smoked cigarettes, packyears and lung dose of inhaled carcinogens

On the example of smoking, the *number of cigarettes smoked per day* is a measurable entity to describe the concentration of carcinogens released by smoking (see 1.1.1). Since also cigarette smoking is time dependent, the number of cigarettes smoked per day during the different phases of smoking covering a time period T were assessed in the German studies,

$$c(t_i) = CD_i,$$

with CD_i denoting the number of cigarettes smoked per day during the ith phase, i=1, ..n.

The exposure accumulated during phase i is then given by *packyears*, PY_i, that is the number of packs (=20 cigarettes) per day times the years of smoking in the ith phase, T_i, with

$T = \sum_i T_i$. A proxy for the cumulative exposure to carcinogens released by smoking during the time period T is thus given by

$$X(T)_{proxy} = PY(T) = \sum_i PY_i = (\sum_i CD_i \cdot T_i) / 20$$

The quantities CD_i and T_i are subject to uncertainties in the assessment process. With the observed quantities denoted by CD_i^*, T_i^*, the derived proxy for the exposure to carcinogens released by smoking („observed" exposure) is $X(T)^*_{proxy} = (\sum_i CD_i^* T_i^*) / 20$.

The individual's *lung dose of inhaled carcinogens released by smoking* differs by certain factors, $l_1,...,l_q$, between persons experiencing the same exposure to carcinogens released by smoking

$$D(T) = l_1 \cdot ... l_q \cdot X(T),$$

with the factors describing between-person-differences in inhaling pattern and bronchial geometry.

3.1.1.4 The operationally defined true predictor

Generally, the true predictor of interest is given by the epidemiological objective on the study data. In the German radon studies, it has been the objective to quantify RR of lung cancer due to „residential radon exposure". „True residential radon exposure" is thus the predictor of interest. Since this is still a theoretical quantity, each investigator has to define a predictor, which is measurable and a valid surrogate[1] for the true exposure of interest: *the operationally defined predictor* (Carroll et al. 1988). Technical, ethical and financial constraints make the definition of such an operationally defined predictor necessary in most observational studies. The conflict between what might be best for estimating the effect of the biological activity

[1] A *surrogate* is defined as a variable which is conditionally independent of the disease status given the predictor of interest, that is the surrogate has no information about the disease status other than what is available in the true predictor of interest.

(dose), preferable to estimate the effect of the environmental burdens and intervention procedures (exposure), and what is feasible for measurement (concentration) is often encountered in epidemiology (Armstrong et al, 1989).

Many studies on lung cancer and residential radon exposure use the TWA radon concentration, TWA(T), as operationally defined predictor, which is measurable by assessing the radon concentration in all relevant homes of the participant, where T covers 5-35 years prior to index date. Hereby, appropriate weighting of the concentrations needs to be considered, since exposures before 15 years are said to loose potential to induce lung cancer (et al, 1994a). OR estimate per 100 Bq/m³ increase in radon exposure are frequently computed.

The German studies were analysed based on two predictors: firstly, the radon gas concentration in the current home, RN_1, and secondly, cumulative radon exposure per year, RNCUM(T), experienced in all relevant homes, where T covers 5-15 years prior to index date, a time period, which is homogeneous regarding the potential to cause cancer. For RN_1, the OR estimates are reported per 100 Bq/m³, for RNCUM(T), per 50 Bq/m³. This accounts for the fact that the average occupancy of the homes, which is accounted for by RNCUM(T), is about half of a 24-hour day; RNCUM(T) is on average about half of the corresponding TWA(T).

The adjustment for the confounder smoking is based on the variable „packyears“, PY(T), in the German studies, where T covers the individual's lifetime. PY(T) is thus used as operationally defined predictor for the exposure to carcinogens released by smoking.

However, if the epidemiological objective was to quantify relative lung cancer risk due to the „true alpha dose“, then „true residential radon exposure“ is merely a surrogate for this predictor. Also, it might be preferable to control for potential confounding due to smoking by adjusting for „lung dose from inhaled carcinogens released by smoking“ instead of the crude surrogate PY(T).

Figure 7 summarises this terminology.

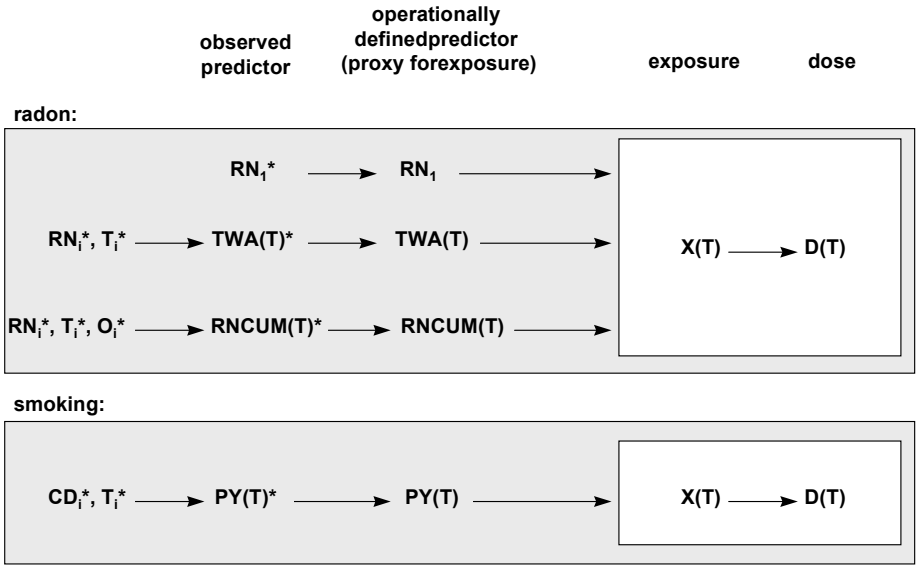

Figure 7: The operationally defined predictor for radon and smoking.

3.1.2 Sources of error

3.1.2.1 Errors in the radon variable

While random error in the exposure assessment from the physical characteristics of the measurement process were studied in great detail (Hardcastle et al. 1996, Lucas et Woodward 1964, Miles 1998, Wrixon et al. 1988), it is rather challenging to investigate the errors in assessing residential radon exposure in the epidemiological setting. Only a very crude list of such errors has been described so far (Bäverstam et al. 1980, Lubin et al. 1995). Here, the components of potential error are summarised systematically. Five stages for estimating residential radon exposure are distinguished:

(1) Estimating average radon gas concentration in the home during exposure of the detector,

(2) Using (1) to estimate average radon gas concentration in the home over the year in which the measurement took place (extrapolation to one year),

(3) Using (2) to estimate radon gas exposure of an individual over a certain period of years prior to the measurement (extrapolation to prior years),

(4) Using (3) (for the current home or for all homes inhabited during a certain

time period, respectively) as a operationally defined predictor for residential radon exposure.

(5) Using (4) as a surrogate for the true alpha dose.

The stages (1-3) describe the deviation between the observed predictor and the operationally defined predictor, the proxy for the exposure, stage (4) the deviation between the operationally defined predictor and exposure, and stage (5) the deviation between the (external) exposure and the (internal) dose. In the following, the error components corresponding to each of the five stages are described. Whether a component is actually relevant depends sometimes on the definition of the operationally defined predictor for residential radon exposure.

(1) Estimating the average radon gas concentration in the home during exposure of detector:

(1.a) Error from *between-measurement-variability* that is the deviation between measurements obtained repeatedly at the same time and place: A measurement by alpha track detectors involves the exposure of a small box of specific geometry containing a thin foil. The emitted alpha particles leave a small trace (track) on the foil. In order to count these tracks, the foil is etched. The specific number of counts of a randomly chosen area of the foil is obtained manually or by a computer program, and the number of counts over the full foil is then extrapolated. The number of radon atoms decayed during the exposure time is derived from the number of tracks on the foil by taking into account that there are three alpha decays in the detector per radon atom (one each from ^{222}Rn, ^{218}Po and ^{214}Po) and simultaneously that the efficiency of finding the tracks by etching is about 30%. Thus, this error component includes

- the error from *background track density* that is the number of tracks observed on a detector which has not been exposed to radon,
- the error from *miscounting the number of tracks* (proportional to the squareroot of the number of tracks).
- the error from *variations in track counting efficiency*: Automatic counters are mostly used. Due to the spectrum of etched track sizes and a background of small tracks which are not from radon or radon daughters, track counters are set to accept a certain range of track sizes as genuine. Any factor that affects the size of etched tracks may change the track counting efficiency. Sources of such variations are the variations in the etching characteristics of the detector material, variations in etching conditions (etchant concentration, temperature, time of etching), variations in counter parameters (illumination and focus), and also ageing and fading effects.
- the error from *calibration* (ideally performed for each batch separately, since the etching differs slightly from batch to batch)
- the error from *underestimating high exposures*, if the exposure is high and the tracks so close together to cause difficulty in distinguishing them after etching.

(1.b) Error from *between-laboratory-variability*: Intercomparison studies have shown a between-lab-variability of about 10% (Kreienbrock et al. 1999).

(1.c) Error from *between-detector-placement-variability*: Radon concentrations deviate between measurements obtained at different locations in the room.

(1.d) Error from *between-room-variability*: Radon concentrations in the rooms without measurements differs from the radon concentration in the living room, which has been used as proxy for the concentrations in the other rooms (except the bedroom).

(1.e) Error from *mis-identifying the house or the detector*: The radon concentration is measured in the wrong house or the false detector is assigned to the participant (For RN_1, the potential of false identification of the house is negligible.).

(2) Extrapolation to one year:

(2.a) Error from *between-season-variability*, if a measurement of less than a year is used to estimate the one-year-average: This error is due to seasonal variations of the radon concentrations in the homes (mainly a response to changes in outdoor temperatures). If seasonal correction is applied, the error from mis-estimating the correction factor still remains.

(3) Extrapolation to prior years:

(3.a) Error from *between-year-variability*: The radon concentration varies from year to year due to differences in atmospheric pressure and temperature.

(3.b) Error from *between-subphase-variability*: The period of time that a house remains unchanged is named a „subphase". The radon concentration during the measurement differs from the concentrations before alterations to the home (building of basements, isolating windows, isolating basements, changing of heating system), which have been shown to affect the radon levels (Gunby et al, 1989). If correction from information on the change of the average radon concentration by certain house alterations is performed (Gerken et al., 1999), the error from mis-specifying the correction factor still remains.

(3.c) Error from *between-owner-variability*: Different ventilation habit of the current owners of the study subject's previous homes leads to conditions in the home during the measurement, which differ from the conditions during the time the study subject has inhabited that home (not for RN_1). If correction from information on the change of the average radon concentration by certain ventilation habit is performed (Gerken et al., 1999), the error from mis-specifying the correction factor still remains.

(4) Using (3) as a proxy for residential radon exposure:

(4.a) Error from the differences in the *ventilation habit depending on room and daytime*: Radon detectors measure average radon concentration for the full day, but the bedroom is occupied during the night and the other rooms during the day. If a participant ventilates the bedroom preferably during the day, the measured concentration in the bedroom underestimates the concentration during the bedroom's occupancy. If a participant sleeps with window open and the window is closed during the day, the measured radon concentration overestimate the concentration during the occupancy. This induces a random error, if it

can be assumed that there is no systematic pattern in the day-night cycle of ventilating the rooms across all study participants (that is some participants sleep with window open, some with window closed).

(4.b) Error from *between-environment-variability*: The radon concentration in residential environments other the principal home (other homes) are usually not measured. The radon concentration in these other environments is assumed to be as high on average as at the prinicipal home. A violation of this assumption would lead to random error, which is handled in this context. Note that it is residential radon exposure that is considered here, which does not include the exposure at the workplace.

(4.c) Error from *between-home-variability*: The radon concentration in the current home is used as proxy for the average radon concentration in all homes inhabited during the time period relevant for the genesis of lung cancer (only for RN_1).

(4.d) Error in *recall of residency time* (for TWA(T) or RNCUM(T)) or *occupancy time* (for RNCUM(T)) of home.

(4.e) Error from *mis-specifying the relevant exposure-window* (not for RN_1): It has been assumed that radon exposure within 5 years before index date cannot be the cause of the lung cancer at index date. Further, exposures from years more than 35 years prior to index date are said to loose their potential to cause cancer.

(4.f) Error from *false recall of relative occupancy of bedroom*: Errors in the recall of the occupancy time of the bedroom induces error in the weighting of the bedroom and living room measurements.

(4.g) Error from *false recall of the absolute occupancy time of home* (only for RNCUM(T)).

(4.h) Error from *ignoring the absolute occupancy time of the home* (for RN_1 and TWA(T)).

(5) Using (4) as proxy for alpha dose:

(5.a) Error in the *equilibrium factor*: An approximate equilibrium factor used for example by Steindorf et al. (1988) is 0.4, which is, however, subject to random uncertainty due the influence of the environmental conditions (temperature, compression).

(5.b) Error from *between-person-variability*: The lung dose of persons with exactly the same residential radon and radon progeny exposure differs between the persons due to

- *Respiratory differences* (inhalation depth and frequency): The amount of radon and radon progeny atoms that are inhaled and enter the bronchial system varies between persons.
- Differences of the *bronchial geometry*: The amount of radon and radon progeny atoms that arrive in the lung and the time that they stay there varies between persons.

Table 2 summarises the components of error in the radon variable. Further, it is indicated, which components were avoided or minimised in the German radon studies.

Table 2: Summary of components of error in assessing radon exposure or alpha dose. The relevance of the component for the specific operationally defined predictor, RN_I, TWA(T), or RNCUM(T) is indicated by x. *In parentheses, „not G" notes that this component is not relevant for the German radon studies, „less G" that this error is reduced in the German radon studies, and „small" that this error is smaller than the ones in the same row.*

Error component	*applicable for*		
	RN-CURRENT	*TWA-RN*	*RNCUM*
(1) Estimating average radon gas concentration in home during exposure of detector			
(1.a) error from between-measurement variability	x	x	x
(1.b) error from between-laboratory-variability	x (not G)	x	x (not G)
(1.c) error from between-placement-variability	x	x	x
(1.d) error from between-room-variability	x	x	x
(1.e) error from mis-identifying the house or the detector	x (small)	x	x
(2) Using (1) to estimate average radon gas concentration in home over the year of the measurement			
(2.a) error from between-season-variability	x (not G)	x	x (not G)
(3) Using (2) to estimate radon gas exposure of an individual over a certain period of years prior to measurement			
(3.a) error from between-year-variability	x	x	x
(3.b) error from between-subphase-variability	x (less G)	x	x (less G)
(3.c) error from between-owner-variability	-	x	x (less G)
(4) Using (3) as operationally defined predictor for residential radon exposure			
(4.a) error from differences in ventilation habit depending on room and daytime	x	x	x
(4.b) error from between-environment-variability	x (less G)	x	x (less G)
(4.c) error from between-home-variability	x	-	-
(4.d) error in recall of residency time	-	x	x
(4.e) error from mis-specifying the relevant exposure-window	-	x	x
(4.f) error from false recall of relative occupancy of bedroom	x	x	x
(4.g) error from false recall of absolute occupancy of home	-	-	x
(4.h) error from ignoring the absolute occupancy time of home	x	x	-
(5) Using (4) as proxy for alpha dose			
(5.a) error in the equilibrium factor	x	x	x
(5.b) error from between-person-variability	x	x	x

Comments to *Table 2*:

- Error from between-measurement-variability was minimised by strict calibration procedures in the lab. Particularly the error from underestimating high exposures was avoided by conducting a short-term measurement to estimate the number of tracks over one year and by initiating multiple consecutive measurements if necessary.
- Error from between-laboratory-variability was avoided, since a single laboratory has conducted all measurements.
- Error from mis-identification of the house was reduced by collecting questionnaire information on the characteristics of the inhabited homes and by cross-checking these with the characteristics of the house encountered at the reported address.
- Error from between-season-variability was avoided by using measurements covering 12 ± 2 months.
- Error from between-owner-variability was reduced by computing a correction factor for the effect of ventilation differences (Gerken et al, 2000), but some error remains by applying one correction factor to a group of persons (those with same ventilation habit) and due to errors in the correction factor estimation (not for RN_1).
- Error from between-subphase-variability was reduced by computing a correction factor for the effect of certain house alterations on the radon concentration (Gerken et al, 2000), but some error remains by applying one correction factor to a group of persons (those with the same house alterations) and due to uncertainties in the correction factor estimation.
- Error from between-environment-variability was lessened by including only subjects with home occupancy of at least 25%.
- Error from ignoring the absolute occupancy of the home is avoided using RNCUM(T), but a pay-off is made by additional error from assessing the occupancy time.

Note that there is no error from *imputing missing radon measurements* in the German studies, since only subjects with available radon measurements were included into the analysis.

3.1.2.2 Errors in the smoking variable

The five stages of uncertainties in assessing the potential confounder, current cigarette smoking, are:

(1) Assessing the number of cigarettes smoked per day.

(1.a) Error from *imprecise recall* of number of cigarettes.

(2) Using (1) to estimate the average number of cigarettes smoked per day during a lifetime.

(2.a) Error from *extrapolating current smoking behaviour to previous behaviour.*

(3) Using (2) to derive the packyears.

(3.a) Error from *imprecise recall of number of years of smoking* by the study participant.

(4) Using (3) as the operationally defined predictor for true exposure to smoking carcinogens.

(4.a) Error from *mis-specifying the relevant exposure-window*: Smoking is assumed as relevant for a certain time period, for example from start of smoking up to 2 years prior to index date.

(4.b) error from *between-brand-variability* that is the variation of carcinogens released by smoking of cigarettes of different brands.

(5) Using (4) as surrogate for the individually experienced lung dose:

(5.a) Error from *between-person-variability* that is differences between persons regarding the depth and frequency of inhaling and the bronchial geometry.

Table 3 summarises the components of error in the smoking variable. The components minimised in the German radon studies are indicated.

Table 3: Summary of components of error in assessing smoking. The applicability of the component in the German radon studies is indicated by x,, „less G" notes that this error was reduced, „not G" denotes that the error was avoided in the German radon studies.

Error component	*applicable*
(1) Estimating the number of cigarettes smoked per day.	
(1.a) false recall of the number of cigarettes smoked per day	less G
(2) Using (1) to estimate the average number of cigarettes smoked per day during a lifetime.	
(2.a) error from extrapolating current smoking habit to previous smoking habit	not G
(3) Using (2) to derive the packyears	
(3.a) false recall of number of years of smoking	less G
(4) Using (3) as operationally defined predictor for lung dose of carcinogens	
(3.a) error from mis-specifying the relevant exposure-window	x
(3.b) error from between-brand-variability	x
(5) Using (5) as proxy for the individually experienced lung dose	
(5.a) error from between-person-variability	x

Comments to *Table 5*:

- Error from extrapolation current smoking behaviour to previous behaviour was avoided by asking for the smoking behaviour during phases covering a lifetime.
- Error of imprecise recall of number of cigarettes or the number of years of smoking was minimised by conducting in-person interviews (no proxy respondent!) by trained staff via detailed standardised questionnaires.

3.1.3 Error structure

3.1.3.1 Classical versus Berkson error

To evaluate the impact of errors in the predictors on the estimate of relative lung cancer risk from radon exposure, it is crucial to classify the error components into errors of classical or Berkson type. Assisting in the differentiation between these types is that, for classical error, an observation repeated under the same condition (i.e. assuming all other errors as non-existing) only allowing for the error under consideration differs from the original observation; on the other hand, for Berkson error, an observation repeated analogously does not change. Also helpful is to understand that, if one measurement is used as a proxy for the average of several measurements, it is a classical error (a repetition of the observation would yield different observations, which vary around the average); on the other hand, if one observation is assigned to a group of persons, it is a Berkson error (a repetition would yield the same result, the true values for each person vary around the assigned observation).

3.1.3.1.1 Errors in the radon variable

The components of error in assessing the radon variable (as identified in Section 3.1.2.2) are classified as follows:

(1.a-c) The error from between-measurement-variability is of the **classical** type: If the measurement was repeated in the same time and place, the observed value would be different and vary about the true value. The error can be assumed as random due to appropriate calibration of the measuring process. The same applies for the error from between-laboratory-variability and the error from between-placement-variability.

(1.d) The error from between-room-variability involves the use of one measurement as a proxy for the average of several measurements: The measurement in the living room is used as proxy for the average of all rooms other

than the bedroom. As noted above, such an error is of the **classical** type.

(1.e) The error from mis-identification of the house or the detector is of the **classical** type: If the measurement is repeated again for a newly specified house/detector, the value would differ from the previous one (It may or may not be the true house/detector now.).

(2.a) The error from between-season-variability is of the **classical** type: If repeated covering a different period of the year, the value would differ. If seasonal correction is applied, this error is replaced by a **Berkson** error due to assigning a certain factor to a group of persons and a **classical** error.due to errors in the correction factor estimation

(3.a) The error from between-year-variability is of the **classical** type: Here, the analogous reasoning as for (1.d) applies, since the use of one measurement (the measured year) as proxy for an average (over all years during a certain time period) is used. No trend over time is assumed, otherwise the error would not be random.

(3.b) The error from between-subphase-variability is of **classical** type: The radon concentration in another subphase (same home, assuming no between-year-difference, etc.) would differ from the original value. If correction for house alterations is applied, this error is replaced by a **Berkson** error due assigning a certain correction factor to a group of persons and a **classical** error due to errors in the correction factor.

(3.c) The error from between-owner-variability is of **classical** type: If the measurement is repeated in the same house but with a different owner, the observed value would differ. If a correction for ventilation habit is applied, this error is replaced by a **Berkson** error due assigning a certain correction factor to a group of persons and a **classical** error due to errors in the correction factor.

(4.a) The error from differences in ventilation habit depending on room and daytime is of **classical** type: If repeated under the same condition but with different ventilation habit, the observed radon concentration differs from the original value.

(4.b-c) The error from between-environment-variability is of **classical** type: With the same reasoning as in (1.d) and (3.a) - the measurement in the principal home is used as proxy for the average over all environments. The analogy applies for the error from between-home-variability.

(4.d) The error from false recall of residency time is of **classical** type: If repeated, the observation would differ from the original.

(4.e) The error from mis-specifying the relevant exposure-window is of the **classical** type: If the exposure-window is redefined, the observed TWA(T) or RNCUM(T) would possibly differ.

(4.f-g) Error in the recall of the relative occupancy time is of **classical** type: If repeated, the observed values would vary about the true occupancy time (assuming no systematic error). The same applies for false recall of the absolute occupancy time.

(4.h) The error from ignoring the absolute occupancy time of home is of **Berkson** type: If the observation is repeated, the same value is observed again.

(5.a) The error from uncertainties in the equilibrium factor is of **Berkson** type: An observation made in another equilibrium state (same home, same year, but different temperature and compression) would provide the same observation of radon exposure as the original observation.

(5.b) The error from between-person-variability is of the **Berkson** type: An observation made for another person

(same home, same residential history, etc.) would provide the same observation of radon exposure as before.

In summary, the components (1.a) to (4.e) are of the classical type, and the components (4.f) to (5.b) are of the Berkson type (with exception of the Berkson type component introduced by applying correction factors). Except for (4.f), the Berkson type is only involved if not the true residential radon exposure, but the true alpha dose is defined as predictor of primary interest. The definition of the predictor hence influences, which of the error components are applicable. For simplicity, the „error in radon exposure“ refers to the summary of the components (1.a) to (4.e), thus a classical error. The „error in using radon exposure as a surrogate for alpha dose“ refers to the summary of the components (4.f) to (5.b), a Berkson error (even if (4.f) is strictly speaking an error in radon exposure). An overview of the classification of the components is given in *Figure 8*.

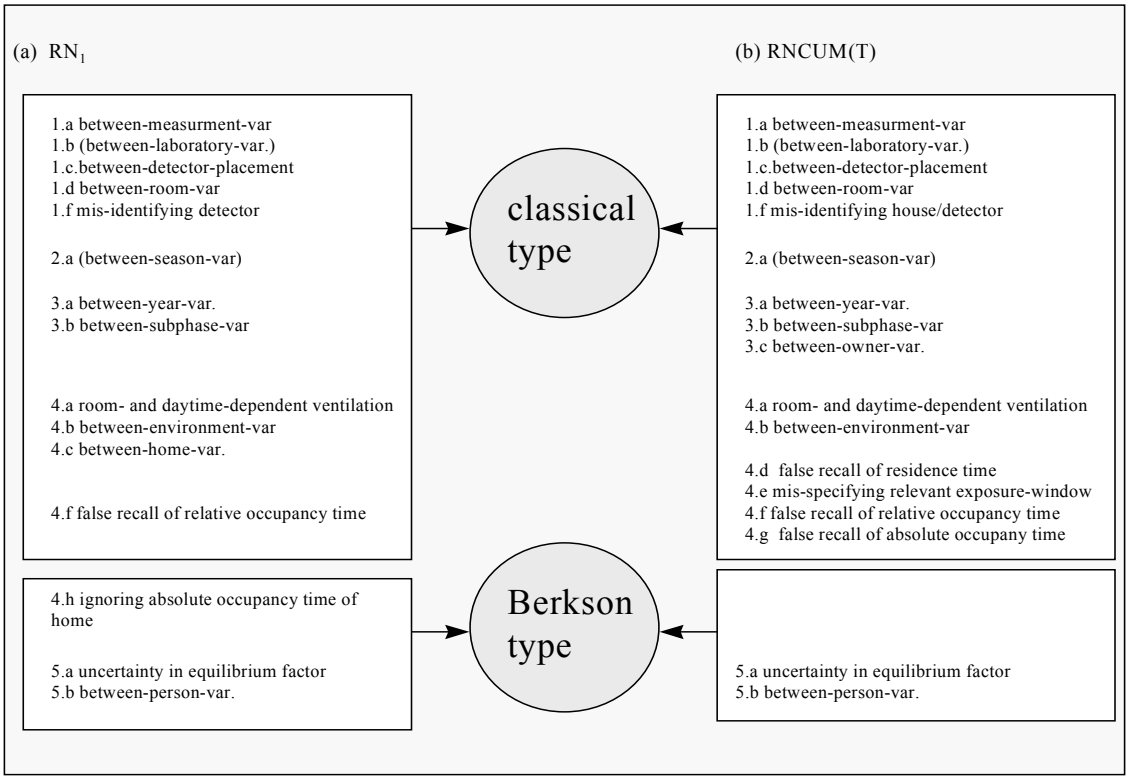

Figure 8: Classification of error components depending on which operationally defined predictor is used, (a) radon concentration in current home (RN_1) and (b) cumulative radon exposure in homes inhabited 5-15 years prior to index date (RNCUM). Components in parentheses are theoretically relevant, but not in the German radon studies.

3.1.3.1.2 Errors in the smoking variable

The components of error in assessing the smoking variable (as identified in Section 3.1.2.2) are classified as follows:

(1.a) The error from imprecise recall of the number of cigarettes smoked per day is of the **classical** type: If the participant is asked again for this information, the number would probably differ, however, it may not vary around the truth, since a participant may systematically over-or underreport. Still, the error over all participants is random, as long as there is only systematic within-person-error, but the systematic error is of different size and direction across all participants. The randomness of the error is violated, if all participants tend to under- or overreport.

(2.a) The error from extrapolating current smoking habit to previous habit is of **classical** type, with the same reasoning as for the extrapolation of the radon concentration to prior years.

(3.a) The error from imprecise recall of the number of years of smoking is of **classical** type, with the same reasoning as in (1.a).

(4.a) The error from mis-specifying the exposure-window is of **classical** type: Same reasoning as for (4.f) regarding the errors in radon exposure.

(4.b) The error from between-brand-variability is of **Berkson** type: Two persons smoking the same number of cigarettes for the same length of time, but smoking different brands, are assigned the same packyears, but e.g. the condensate concentration is very likely to vary between brands.

(5.a) The error from between-person-variability is of **Berkson** type: Same reasoning as for (4.b).

For simplicity, the components (1a), (2a), (3.a) and (4.a) summarise the „error in packyears“, a the classical error. The components (4.b) and (5.a) summarise „the error from using packyears as a surrogate for the lung dose of inhaled carcinogens released by smoking“, a Berkson error. The *Figure 9* provides an overview.

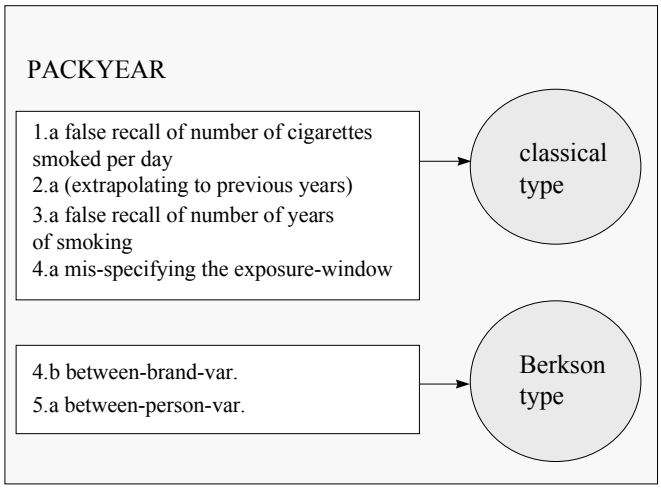

Figure 9: Overview: Classification of error components for the smoking variable, if packyears are the operationally defined predictor for current cigarette smoking. Components in parentheses are theoretically relevant, but not in the German radon studies.

3.1.3.2 Multiplicativity versus additivity of error

Repeated measurements of radon concentration in several homes over a sequence of years showed a multiplicative structure (Lubin et al., 1995, Darby et al., 1998, Lagarde et al., 1997) and extensive physical experiments have identified the factors that influence radon concentrations (Gunby et al., 1997, Hardcastle et al., 1985, Wrixon et al., 1995, Miles et al., 1992). The multiplicativity is plausible from the fact that the modifiers of the radon concentration such as temperature or the ventilation always affect a proportion of the radon atoms in the room. Therefore, the multiplicative model for the classical error in radon exposure and for the Berkson error in using radon exposure as surrogate for the alpha dose is appropriate.

It can be reasoned that a multiplicative error is also applicable for the classical error in packyears and the Berkson error in using packyears as surrogate for inhaled lung dose of carcinogens released by smoking: This error is likely to be almost zero for non-smokers, small for moderate smokers and larger for heavy smokers. The multiplicative error structure implies such a constantly increasing deviation between true and assessed predictor with increasing

levels of the predictor (see Section 2.1.1.3). In fact, the OR estimates were adjusted for the packyears+1 on the original scale, instead of log(packyears+1) as in the primary analysis, to ensure a multiplicative error model for both predictors, radon and smoking. If the OR estimates had been adjusted for packyears on the log-scale, a mixed error model - a multiplicative error in radon and an additive error in the smoking - would have been required. The regression calibration allows for such modelling in a quite simple way, but only if the correlation between the predictors is ignored. If the correlation is to be taken into account - and the investigation of the impact of this correlation on the error correction is one of our declared objectives - a mixed error model represents a severe obstacle and is thus not considered here.

3.1.3.3 Random versus systematic error

All errors are assumed to be random . For example, the error of false recall of number of cigarettes smoked per day would be systematic, if all study participants under-reported their smoking behaviour. Also, the error from between-year-variability in radon exposure would be systematic, if the radon concentrations over all houses showed a trend over time (e.g. due to climate change).

3.1.3.4 Homoscedastic versus heteroscedastic error

All errors are assumed to be homoscedastic. Heteroscedastic error would occur for example, if radon concentrations for homes inhabited a long time before index date were prone to a larger error than for homes inhabited more recently (due to the higher probability of false identification of the house or more alterations to the house). A heteroscedastic error could also arise, if the recall of smoking habit was less reliable for smoking in the past.

3.1.4 Error size

3.1.4.1 Some measures of error size

The error size is usually given as the SD of the error on the original scale for additive error, σ_E, and on the log-scale for multiplicative error, $\sigma_{\log E}$.

For multiplicative error, alternatives are: the GSD of the error (Lubin et al. 1995), $\exp(\sigma_{\log E})$, the coefficient of variation (CV) as the error's SD on the original scale divided by the mean on the original scale (Darby et al. 1998), $\sqrt{\exp(\sigma^2_{\log E}) - 1}$. For small $\sigma_{\log E}$, this CV is close to $\sigma_{\log E}$ (since then $\exp(\sigma^2_{\log E})$ is close to $\sigma^2_{\log E} + 1$). Also reported is a CV interpreted as the SD on the original scale divided by the GM (Lagarde et al. ,1997, Lubin et al., 1995), $\exp(\sigma_E{}^2) \cdot \sqrt{\exp(\sigma_E{}^2) - 1}$. For small $\sigma_{\log E}$, the two definitions of the CV provide similar values. The following table is a conversion table for these measures.

Table 4: Some measures of error size (for multiplicative error), with SD denoting standard deviation, GSD denoting geometric standard deviation and CV denoting coefficient of variation.

SD of log of error	*Variance of log of error*	*GSD of error*	*CV as SD divided by mean*	*CV as SD divided by GM*
0.2	0.04	1.22	20%	20%
0.3	0.09	1.35	31%	32%
0.4	0.16	1.49	42%	45%
0.5	0.25	1.65	53%	60%
0.6	0.36	1.82	66%	79%
0.7	0.49	2.01	80%	102%
0.8	0.64	2-23	95%	131%
0.9	0.81	2.46	112%	268%
1.0	1.00	2.72	131%	216%

3.1.4.2 Proportion of error variance to the variance of the observed predictor

Since the measures of error size given above are not very intuitive, it is helpful to state the error variance as proportion of the variance of the observed predictor (on original scale for additive error, on log-scale for multiplicative error). For classical error, this proportion is given by the reliability factor. It describes the proportion of the observed predictor variance which is explained by the error and which would disappear, if the predictor was measured without error. .For Berkon error, this proportion provides the factor, by which the variance of the true predictor exceeds the variance of the observed predictor and by which the variance of the observed predictor would increase, if the error was eliminated.

Table 7 summarises the SD of the error in radon and smoking among controls or cases, so that the error variance meets a certain percentage of the observed predictor variance (on log-scale for multiplicative error, on original scale for additive error). For example, the variance of an error in radon exposure with SD of 0.5 is 75% of the observed radon exposure variance among West controls (on log-scale). The variance of an error of that size in smoking is 10% of the variance of observed packyears among West controls (on the log-scale).

A classical error of 75% explains 75% of the observed predictor variance, a Berkson error of 75% indicates that the true predictor variance is 1.75 times as large as the observed predictor variance. Note that the classical type error cannot exceed 100%, since the variance of the classical error is limited by the observed predictor variance (The most extreme situation is one where the error explains 100% of the observed predictor variance and thus the true predictor variance would not spread at all: The true predictor is just one given value.). On the other hand, the Berkson error can increase, theoretically, indefinitely.

In the sensitivity analysis exploring the impact of differently sized classical error in the risk factors on the OR estimate in the German radon studies, errors of up to 50% are considered. Reference distribution is the distribution of radon exposure observed among controls of the West study and the distribution of packyears+1 observed among controls of the East study. Exploring the impact of Berkson error, errors of up to 100% are considered.

Table 5: German radon studies: Table entries are the SD of the error for radon and smoking (packyears+1), so that the error variance meets a given percentage of the variance of radon exposure or packyear+1 (on log-scale for multiplicative error, on original scale for additive error) observed among controls or cases. Highlighted are the error sizes that are of special interest.

% of variance of (log of) observed exposure		*Standard deviation of (log of) error*		
		German West controls / cases	*German West high controls / cases*	*German East controls / cases*
multiplicative error				
radon:	300%	1.0 / 1.0	1.0 / 1.0	1.1 / 1.2
	100%	0.6 / 0.6	0.6 / 0.6	0.6 / 0.7
	75%	0.5 / 0.5	0.5 / 0.5	0.5 / 0.6
	50%	0.4 / 0.4	0.4 / 0.4	0.4 / 0.5
	25%	0.3 / 0.3	0.3 / 0.3	0.3 / 0.3
	10%	0.2 / 0.2	0.2 / 0.2	0.2 / 0.2
smoking :	300%	2.7 / 1.9	2.6 / 1.9	2.5 / 2.0
	100%	1.5 / 1.1	1.5 / 1.1	1.4 / 1.2
	75%	1.3 / 1.0	1.3 / 0.9	1.2 / 1.0
	50%	1.1 / 0.8	1.1 / 0.8	1.0 / 0.8
	25%	0.8 / 0.5	0.8 / 0.5	0.7 / 0.6
	10%	0.5 / 0.3	0.5 / 0.3	0.5 / 0.4
additive error				
radon:	300%	81 / 67	114 / 83	81 / 67
	100%	47 / 39	66 / 48	47 / 39
	75%	40 / 33	57 / 41	40 / 33
	50%	33 / 27	46 / 34	33 / 27
	25%	23 / 19	32 / 23	23 / 19
	10%	15 / 12	21 / 15	34 / 37
smoking:	300%	34 / 37	31 / 34	23 / 28
	100%	20 / 21	18 / 20	13 / 16
	75%	17 / 18	15 / 17	12 / 14
	50%	14 / 15	13 / 14	10 / 11
	25%	10 / 10	9 / 10	7 / 8
	10%	6 / 7	6 / 6	4 / 5

3.1.4.3 Range of observed predictor values given a certain true value

A concrete view on the error size is provided by presenting the range of the predictor values that would possibly be observed given a true predictor value.

For a multiplicative classical error, the variance of the log of the observed predictor, Z, given a true predictor value of x, $\text{var}(\log Z|x)$, is given by the variance of the log of the error, $\sigma^2_{\log EM}$; the average of the observed values is given by $\log x$. Assuming that the error and the true predictor are lognormally distributed , the distribution of $\log Z|x$ normal. Therefore, 95% of the log of the observed values are known to lie within $[\log x - 2\sigma_{\log EM}, \log x + 2\sigma_{\log EM}]$, and 95% of the observed values on the original scale lie within

$$[x \cdot \frac{1}{(\text{GSD}[EM])^2}, x \cdot (\text{GSD}[EM])^2],$$

where $\text{GSD}[EM]$ denotes the GSD of the error.

For an additive classical error, the variance of the observed predictor given a true predictor value of x, $\text{var}(Z|x)$, is given by the variance of the error, σ^2_{EA}; the average of the observed values is given by x. Assuming that the error and the true predictor is normally distributed, the distribution of $Z|x$ is normal and 95% of the observed values are known to lie within

$$[x - 2\sigma_{EA}, x + 2\sigma_{EA}].$$

Table 6 displays the range of 95% of the observed radon exposures that would be observed if a certain true radon exposure was given.

Table 6: Error size: Range of 95% of the observed radon exposures given a certain true radon exposure. The error size is given as the SD of the error (on original scale for additive, on log-scale for multiplicative error) and in parentheses as the percentage of the error variance when compared to the observed radon exposure variance (on original scale for additive, on log-scale for multiplicative error).

	Error size	*True radon exposure in Bq/m³*	*Range of 95% of observed radon exposure in Bq/m³*
multiplicative error	0.3 (25%)	5	3 to 9
		50	27 to 91
		500	270 to 910
		5000	2700 to 9100
	0.4 (50%)	5	2 to 11
		50	22 to 111
		500	220 to 1110
		5000	2200 to 11100
	0.5 (75%)	5	2 to 14
		50	18 to 136
		500	184 to 1360
		5000	1800 to 13600
additive error	20 (25%)	5	-35 to 45
		50	10 to 90
		500	460 to 540
		5000	4960 to 5040
	30 (50%)	5	-55 to 65
		50	-10 to 110
		500	440 to 560
		5000	4940 to 5060
	40 (75%)	5	- 75 to 85
		50	- 30 to 130
		500	420 to 580
		5000	4920 to 5080

The analogous procedure is applied for the smoking variable, packyears+1. For simplicity, let us assume that the smokers have smoked for 20 years. Then the packyears equals the number of cigarettes smoked per day, which is easier to interpret. Note that the smoking variable enters the exposure-disease-model as packyears+1 and that therefore non-smokers enter the model with value 1. Due to the fact that the multiplicative error model involves a zero error, when the true predictor is zero, there is a small error involved even for non-smokers with a predictor value of 1. This does not seem to be very realistic: A true non-smoker would

probably not recall having smoked regularly a non-zero number of cigarettes per day. However, it can be seen in the table that the range of observed values is very small and can be neglected. Further, data on self-reported smoking behaviour and measured serum cotinine levels showed that 7.5% of the self-reported smokers have serum cotinine levels associated with a non-smoking status (Caraballo et al., 2001).

It can be further seen from *Table 7*, that the range of reported number of cigarettes b becomes unrealistic for large error: For example, for an error size of 1.0, that is an error explaining about 50% of log of the observed packyear variance in the East study, a person having smoked 2 pack a day regularly for 20 years would possibly report smoking 5 cigarettes per day (A person smoking so much would at least report 1 pack a day.) or up to over 7 pack (that means smoking about 10 cigarettes per hour!). A person having smoked 1 cigarette per day would report up to 1 pack a day (completely unrealistic!).

Heuristically, an error of size 0.5 can be motivated that is an error explaining 10% of the observed packyear variance in the West study (12% in East): This implies that participants, who have smoked one cigarette per day for 20 years (1 packyear), might report anything between 0 to 4 cigarettes per day; those having smoked 10 cigarettes per day for 20 years (10 packyears) report 3 to 29 cigarettes; participants having smoked 1 pack a day for 20 years (20 packyears) report 7 to 56 cigarettes per day, and those having smoked 2 packs a day for 20 years (40 packyears) might report 1 pack up to 5 packs a day. This seems to be plausible. Still, the error in the reported number of years of smoking is not yet accounted for. But in order to stay conservative, an error of size 0.5 is assumed to be most realistic. However, error sizes of up to 1.0 are explored to provide comparability with the impact of the error in radon exposure.

Table 7: Error sizes: The analogy to Table 6 for packyears.

	Error size	*True # of cigarettes (packs) smoked per day for 20 years*	*95% of reported # of cigarettes (packs) smoked per day for 20 years*	
multiplicative error	0.5 (10%)	0	0 to 2	
		1	0 to 4	
		10 (0.5)	3 to 29	(0.2 to 1.5)
		20 (1)	7 to 56	(0.4 to 2.8)
		40 (2)	14 to 110	(0.7 to 5.5)
	0.7 (25%)	0	0 to 3	
		1	0 to 7	
		10 (0.5)	2 to 44	(0.1 to 2.2)
		20 (1)	4 to 84	(0.2 to 4.2)
		40 (2)	9 to 165	(0.5 to 8.3)
	1.0 (50%)	0	0 to 6	
		1	0 to 14	
		10 (0.5)	0 to 80	(0 to 4)
		20 (1)	2 to 154	(0.1 to 7.7)
		40 (2)	5 to 302	(0.3 to 15.1)
	1.2 (75%)	0	0 to 10	
		1	0 to 21	
		10 (0.5)	0 to 120	(0 to 6)
		20 (1)	1 to 230	(0.1 to 11.5)
		40 (2)	3 to 451	(0 to 23)
additive error	4	0	-8 to 8	
		1	-7 to 9	
		10	2 to 18	
		20	12 to 28	(0.6 to 1.4)
		40	32 to 48	(1.6 to 2.4)
	7	0	-14 to 14	
		1	-13 to 15	
		10	-4 to 24	
		20	6 to 34	(0.3 to 1.7)
		40	26 to 54	(1.3 to 2.7)
	10	0	-20 to 20	
		1	-19 to 21	
		10	-10 to 30	
		20	0 to 40	(0 to 2)
		40	20 to 60	(1 to 3)
	12			
		0	-24 to 24	
		1	-23 to 25	
		10	-14 to 34	
		20	-4 to 44	(- 0.2 to 2.2)
		40	16 to 64	(0.8 to 3.2)

3.1.5 Information from replicate data

By means of measurements repeated in the homes and over time, further prove of the multiplicativity or the error and complementary information on error size is provided.

3.1.5.1 Bedroom and living room measurements

Information on the between-measurement-variability under epidemiological conditions is provided by the bedroom and living room measurements - given the between-room-differences can be controlled for. For houses inhabited by the controls of the West-High study at index date with bedroom and living room at the same floor, *Figure 10* shows all measurements in the rooms versus the mean of the measurements by house (The picture for the West or the East study would look very much alike, but with three times as much observations.). It can be seen that the vertical spread of measurements increases with increasing mean concentration in (a), whereas in (b) the spread is rather constant: The graph shows the „trumpet“ on the original scale and the „tube“ on the log-scale, which indicates a multiplicative structure of the error from between-measurement-variability (see 2.1.1.3).

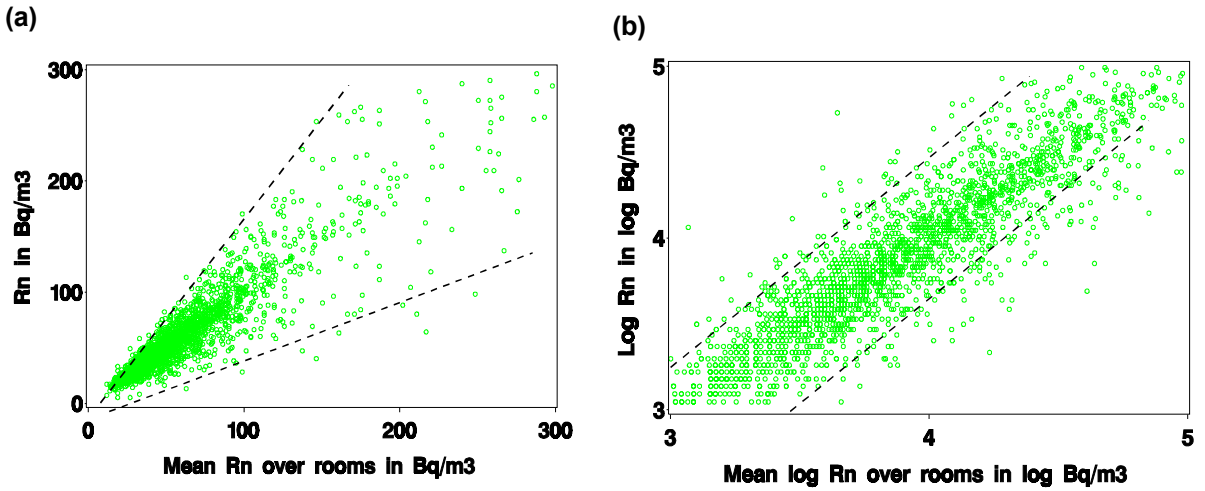

Figure 10: Bedroom/living room measurements with both rooms at the same floor (East study): Radon concentrations for each house in each room versus the mean radon concentration by house (a) on the original scale and (b) on the log-scale.

The results of the analysis, based on the log of the measurements, are summarised in *Table 8*. The estimated size of the error from between-measurement-variability is about 0.3 for all studies. The effect of the room (bedroom versus living room) is about 10% in the West and West-High study and 30% in the East study. Additionally modelling a fixed effect of the floor does not influence the estimated error size, but reduces the room effect (to 5% in West and West-High, to 20% in East).

Table 8: Bedroom/living room measurements: For each method of analysis, the estimated error size (that is the SD of the log of the error) is given for the error from between-measurement variability, Err(meas). Further, the estimates of the fixed effect of the room (B= bedroom, L=living room) and of the floor level are displayed.

study	*method of analysis*	*error size*	*room effect*		*floor effect*		
		Err(meas)	*B*	*L*	*0*	*1*	*>1*
West	$\log Z_{i,j,k} = \mu + HOME_i + ROOM_j + \varepsilon_{i,j,k}$	0.28	1.0	1.10	-	-	-
	$\log Z_{i,j,k} = \mu + HOME_i + ROOM_j + FLOOR_j + \varepsilon_{i,j,k}$	0.27	1.0	1.05	1.41	1.20	1.0
West-High	$\log Z_{i,j,k} = \mu + HOME_i + ROOM_j + \varepsilon_{i,j,k}$	0.27	1.0	1.12	-	-	-
	$\log Z_{i,j,k} = \mu + HOME_i + ROOM_j + FLOOR_j + \varepsilon_{i,j,k}$	0.27	1.0	1.07	1.42	1.24	1.0
East	$\log Z_{i,j,k} = \mu + HOME_i + ROOM_j + \varepsilon_{i,j,k}$	0.33	1.0	1.29	-	-	-
	$\log Z_{i,j,k} = \mu + HOME_i + ROOM_j + \varepsilon_{i,j,k}$	0.32	1.0	1.22	1.62	1.22	1.0

*: $Z_{i,j,k}$ is the radon concentration measurement in the jth room (0=living room, 1=bedroom) on the kth floor (0= ground floor or basement, 1 = first floor, 2 = second floor or higher) in the ith house.

3.1.5.2 Year-by-year replicates

The year-by-year replicates provide estimates of the error from between-year-variability and from between-measurement-variability. However, the radon concentrations in the houses measured here are extremely high and not at all representative of the radon levels encountered in the homes of the general population. The graphs in *Figure 11* summarise the distribution of the measurement by house, by year and by house and year.

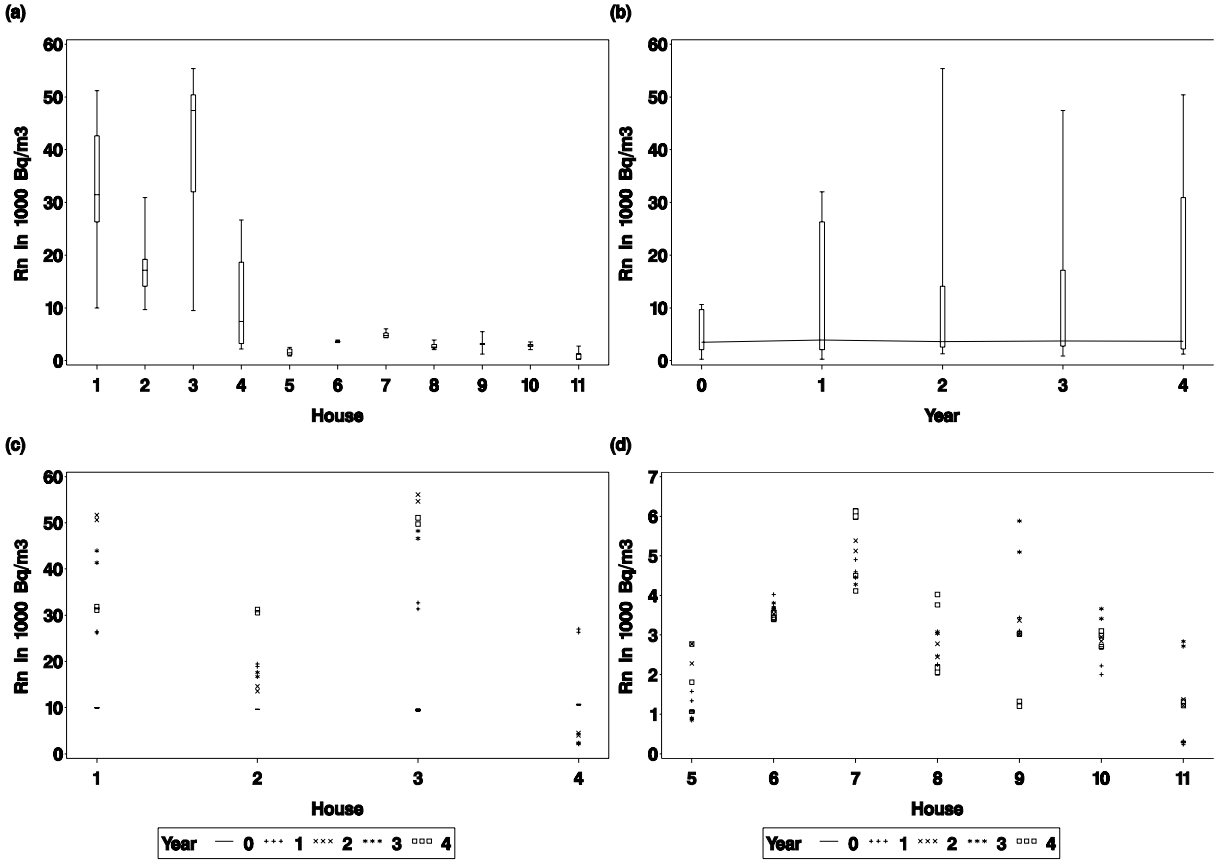

Figure 11: Year-by-year replicates (kindly provided by R. Lehmann): The boxes indicate the inter-quartile-range, the whiskers the full range of the radon concentration measurements (a) by house and (b) by year. The horizontal line in each box indicates the median; in (b), a line joins these medians. The single measurements are displayed by house and year in (c) for the very highly exposure houses 1-4 and in (d) for the less exposed houses 5-11 (Note the different scaling of the vertical axis in (c) and (d)!).

It is obvious from (a) that the houses 1-4 are extremely highly exposed when compared to the houses 5-11. In (b), the line joining the medians of the measurements by year shows no trend over time. The more detailed graphs (c) and (d) with identical symbols for measurements in the same home and year illustrate, that the difference between the measurements in the same home and year and thus the between-measurement-variability is rather small. Further, (c) and (d) shows that the between-year-variability is quite large as compared to the between-measurement-variabilility.

Figure 12 shows all measurements versus the mean of these measurements by house (a) on original scale and (b) on log-scale. It can be seen that the vertical spread of the radon concentrations per house increases with increasing mean in (a), whereas in (b) the spread is rather constant: The graph shows the „trumpet“ on the original scale and the „tube“ on the log-scale, which indicates a rather multiplicative structure of the error from between-year-variability (see Section 2.1.1.3).

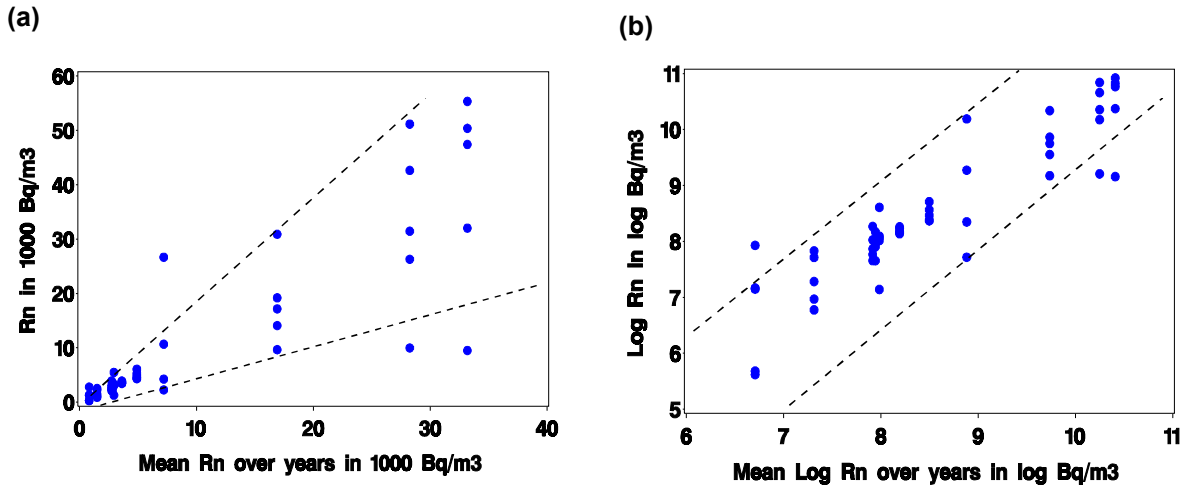

Figure 12: Year-by-year replicates (kindly provided by R. Lehmann): Radon concentration measurements („corrected“ for between-measurement error by taking the mean of the two measurements in the same year and home) versus the mean of these measurements in one house (a) on the original scale and (b) on the log-scale.

The results of the analysis based on the log of the measurements are summarised in *Table 9*. Computing the mean of the variance of replicates provides similar estimates as applying the ANOVA model. The error from between-year-variability is large when compared to the error from between-measurement-variability (about 0.6 versus 0.07), which coincides with our observation viewing the plotted data. It is interesting to note that the estimate of the size of the combined error is smaller than if the error from between-year-variability is considered alone. Not reported are the results from the model including a fixed effect of the years, since the estimates are similar and the fixed effect for years is not statistically different from zero (p-value=0.2), which indicates - as the plotted data in *Figure 11b* - that there is no trend over

time. Removal of each house in turn reveals that the houses 4 and 11 are most influential. The results without one of these (no matter which one), reported in parentheses in the table, show a slightly smaller size of the error from between-year-variability.

Table 9: Year-by-year replicates (kindly provided by R. Lehmann): For each method of analysis, the estimated error size (that is the SD of the log of the error) is given for the error from between-measurement variability, Err(meas), the error from between-year-variability, Err(year), and for the combination of both. The error sizes in parentheses are computed without house 11.

method of analysis	*error size*		
	Err(meas)	*Err(year)*	*both*
$\sqrt{\underset{i}{\mathrm{MEAN}}\left(\underset{j}{\mathrm{VAR}}\left(\underset{k}{\mathrm{MEAN}}(\log Z_{i,j,k})\right)\right)}$ *	-	0.60 (0.54)	-
$\sqrt{\underset{i}{\mathrm{MEAN}}\left(\underset{j,k}{\mathrm{VAR}}(\log Z_{i,j,k})\right)}$	-	-	0.56 (0.51)
$\sqrt{\underset{,ji}{\mathrm{MEAN}}\left(\underset{k}{\mathrm{VAR}}(\log Z_{i,j,k})\right)}$	0.07 (0.07)	-	-
$\log Z_{i,j,k} = \mu + \mathrm{HOME}_i + \mathrm{HOME}_i \cdot \mathrm{YEAR}_j + \varepsilon_{i,j,k}$	0.07 (0.07)	0.58 (0.51)	-
$\log Z_{i,j,k} = \mu + \mathrm{HOME}_i + \varepsilon_{i,j,k}$	-	-	0.55 (0.49)

*: $Z_{i,j,k}$ is the kth radon concentration measure(k=1,2) in the jth year (j=1,..., 5) and the ith house (i=1,..., 11).

3.1.5.3 Intercomparison data

The intercomparison data provides, again, information on the error from between-measurement variability. Compared to the year-by-year replicates, this data has the advantage that the measurements were conducted in houses with exposures typical for the German radon studies. Compared to the bedroom/living room measurements, it has the advantage that there is no need to worry about room-differences. However, this data is sparse including only 5 houses, and the laboratory personnel has been aware of the intercomparison study.

Figure 13, displaying the original data by house, shows that the between-measurement-

variability is small and that the variability is slightly larger for the two houses with larger radon concentrations (houses 3 and 4), which indicates a multiplicative error structure.

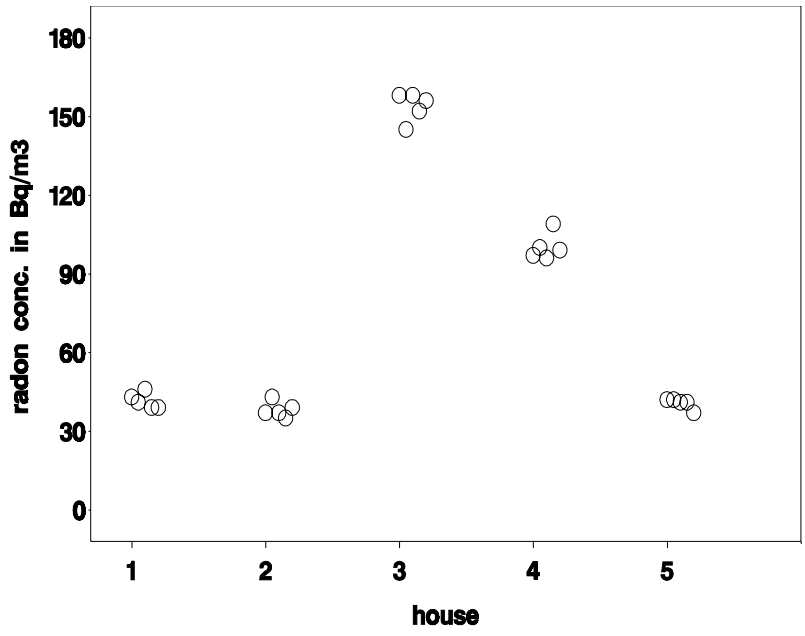

Figure 13: Intercomparison data (kindly provided by L. Kreienbrock): Radon concentration measurements by house.

The estimate of the error from between-measurement-variability, obtained as the squareroot of the mean of the within-house-variances or as the SD of the residual of the ANOVA model, is 0.06.

3.2 Interpreting the correction methods

The regression calibration (RC) method (see Section 2.1.3) provides modified logistic regression models allowing for errors in the predictors. In the following, it is shown that some of the impact of correcting for error in the predictors on the RR estimate can be derived analytically from the model formulation. This enables the differentiation between effects explained by the theory and effects inherent in the data, when applying the error correction on

the German radon studies' data. In 3.2.5, the differences to the approximate likelihood (AL) method, described in 2.1.4, are illustrated. Errors are assumed to be „well-behaved", that is the errors are random, homoscedastic, non-differential towards disease status and uncorrelated.

3.2.1 Errors in the predictor of primary interest without covariate

In the following, the impact of error in the variable of primary interest, Z_1, on the effect parameter estimate, β_{Z_1}, is studied, when no covariate is included into the exposure-disease model (Z_2 set to zero.). Note that a *covariate* is defined in this context as a predictor variable other than the predictor of primary interest, e.g. Z_2.

3.2.1.1 Impact on the effect estimate of primary interest

In the following, the direction of the bias from error in the primary predictor is explored, when no covariate is included into the exposure-disease model.

If no covariate is adjusted for, the exposure-disease model correcting for <u>additive classical error</u> is $\text{Logit}(p) = \tilde{\alpha} + \beta_{T_1}\gamma_1 Z_1$ (from (11) with Z_2 =0) and thus involves entering $\gamma_1 Z_1$ instead of Z_1 into the model. Since γ_1 takes on values between zero and unity, the corrected predictor values are all drawn towards zero. The range of the predictor is thus reduced. Correcting for such error should therefore increase the absolute value of the effect parameter (In other words, the RC method predicts that such error induces bias towards the null hypothesis.).

Figure 14 illustrates, why a reduced range of the predictor generally increases the regression coefficient on the example of linear regression, which is easier to display than the logistic regression. The same applies for the logistic regression coefficient.

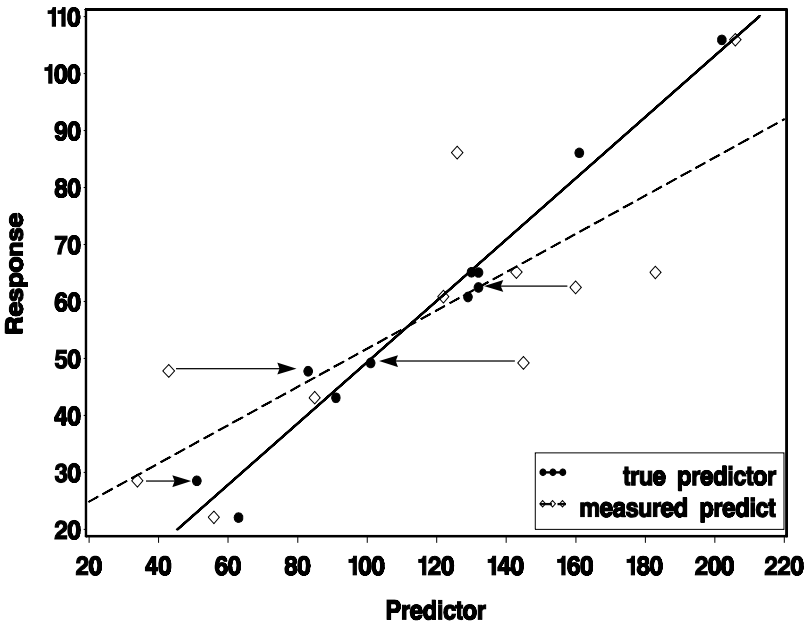

Figure 14: Attenuation of the naive slope (fictitious data): Diamonds indicate the predictor measured with classical error with its response, dots symbolise the true predictor with its response. The hatched line is the linear regression line through the diamonds (naive slope), the solid line is though the dots (true slope).

It can be seen that, for each diamond, the corresponding dot is mostly found shifted horizontally to the right for small predictors and to the left for large predictors: The diamonds are on average shifted towards the mean value of the predictor, and the horizontal range of the diamonds, that is the range of true exposure, is thus smaller than the horizontal range of the dots, the range of measured exposure. Drawing the line regression through the diamonds (hatched line) and another one through the dots (solid line), it can be seen that the slope of the solid line (true slope) is steeper than the slope of the hatched line (naive slope): The slope, that is the regression coefficient, is thus increased, if the regression is based on the true predictor values instead of the measured.

More simply, for homoscedastic additive classical error, the naive beta-coefficient estimate of the effect of primary interest, $\hat{\beta}_{Z_1}$, is given by $\hat{\beta}_{T_1}\gamma_1$. Its absolute value is therefore smaller than the corrected beta-coefficient, $\hat{\beta}_{T_1}$. The bias from such error is thus, independent from

the data, towards the null hypothesis. For example, if the reliability factor, γ_1, equals 0.5, indicating that 50% of the exposure variance is explained by error, the naive beta-coefficient estimate is only half of the true coefficient. Note that „bias towards the null hypothesis" implies that the naive OR estimate is closer to unity than the truth (attenuation): A naive OR estimate above unity is higher after correction for error (The adverse effect increases.); a naive OR below unity is smaller after correction (The protective effect increases.).

If the predictor is measured with <u>additive Berkson error</u>, γ_1 equals unity. Accounting for such error therefore does not change the exposure-disease model; there is no bias on the OR estimate.

The exposure-disease model correcting for <u>multiplicative classical error</u> in the predictor is $\text{Logit}(p) = \alpha + \beta_{T_1} m_1 Z_1^{\gamma_1}$ (from (10) with $Z_2 = 0$) and thus involves entering $m_1 Z_1^{\gamma_1}$ instead of Z_1 with $m_1 = (e^{\mu_1})^{1-\gamma_1} e^{0.5\gamma_1 \sigma^2_{\log EM_1}}$ with $\gamma_1 = 1 - \sigma^2_{\log EM_1} / \sigma^2_{\log Z_1}$. The scaling by m_1 is just a linear transformation of the exposures; the exponentiation by γ_1 is a root function (since $\gamma 1$ takes on values between zero and unity). A root function generally draws the (positive) values towards unity. The transformation of Z_1 into $m_1 Z_1^{\gamma_1}$ therefore draws the exposure towards m_1, which equals the expectation of true exposure exponentiated by $1-\gamma_1$, $(E[X_1])^{1-\gamma_1} = (e^{\mu_1})^{1-\gamma_1} (e^{0.5\sigma^2_{\log X_1}})^{1-\gamma_1}$. Since the extreme values are reduced by the correction, the range of exposure is decreased. With the same reasoning as for the additive classical error, correcting for multiplicative classical error should therefore increase the absolute value of the effect parameter. In other words, such error induces bias towards the null hypothesis. However, the effect of correction can depend on the individual data; the corrected beta-coefficient estimate cannot be directly computed from the naive estimate as in the case of additive classical error.

If the predictor is measured with <u>multiplicative Berkson error</u>, γ_1 equals unity and $m_1 = 0.5 e^{\sigma^2_{EMB_1}}$. The exposure-disease model correcting for such error is $\text{Logit}(p) = \alpha + \beta_{T_1} \cdot e^{0.5\sigma^2_{\log EMB_1}} Z_1$, which involves entering $e^{0.5\sigma^2_{\log EMB_1}} Z_1$ instead of Z_1. This implies that the exposures are scaled by a constant larger than unity (The variance $\sigma^2_{\log EMB_1}$ is larger than zero and the exponential function of positive values results in values

larger than unity.), which implies that the multiplicative Berkson error, as predicted by the RC method, has the opposite effect as the multiplicative or additive classical error: The range of the predictor widens; the absolute value of the effect parameter thus decreases after correction. In other words, such error induces bias away from the null. This conclusion is not data-dependent as for the multiplicative classical error: For homoscedastic multiplicative Berkson error, the naive beta-coefficient estimate, $\hat{\beta}_{Z_1}$, can be directly computed from the corrected estimate as $e^{0.5\sigma^2_{\log EMB1}} \cdot \hat{\beta}_{T_1}$, which is thus larger than $\hat{\beta}_{T_1}$.

Figure 15 shows the exposure-disease-relationship that is observed when the predictor is measured with additive/multiplicative classical/Berkson error. It can be seen that the additive Berkson has no impact, the multiplicative Berkson error induces bias away from the null hypothesis, and the additive and multiplicative classical error induce bias towards the null hypothesis. Note that the bias from multiplicative Berkson error seems to be less severe than the bias from (additive or multiplicative) classical error. It can be further seen that the true exposure-disease-relationship based on a logistic model is nearly linear for this range of the predictor, and the errors, except the multiplicative classical error, do not change the observed exposure-disease-relationship other than by scaling. With the predictor measured with multiplicative classical error, the observed exposure-disease-relationship is shaped differently, rather like a root function than a linear line. The curve is steeper for very small predictor values and less steep for larger values. Therefore, ignoring the multiplicative classical error is equivalent with fitting a logistic model, which can be approximated here by fitting a linear line, to this root-function-shaped curve leading to a severe mis-specification of the exposure-disease-relationship.

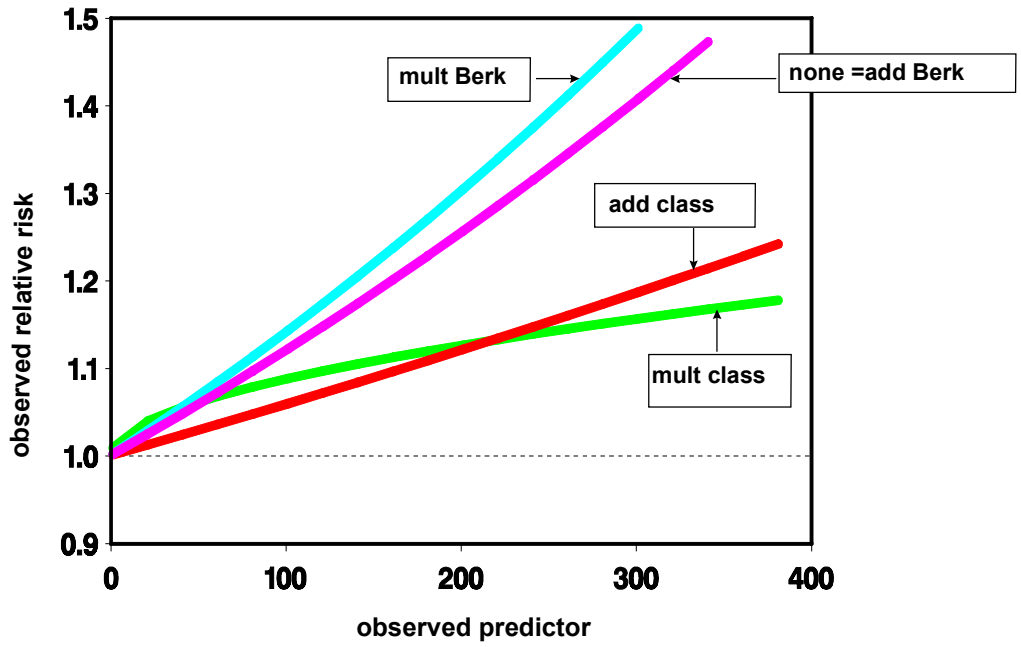

Figure 15: Theoretically observed exposure-disease-relationship when the predictor is measured with errors: additive error of classical or Berkson type (with SD of the error of 50) or multiplicative error of classical or Berkson type (with SD of the log of the error of 0.4). The underlying true exposure-disease-relationship is given by the logistic regression model and a true relative risk of 1.12 (see „none").

3.2.1.2 Dependence of the bias on predictor distribution parameters

Frequently, groups with differing predictor distribution parameters are compared. For example, the estimate of relative lung cancer risk due to radon exposure in West-High, a subgroup from radon-prone areas of the West study, was contrasted with the estimate in the West study in the primary analysis, and this subgroup exhibits larger GM and GSD of radon exposure than the full data. Further, in the meta-analysis of several studies on the same research topic, the predictor distribution parameters differ often quite notably: For example, the Finish, the English, the German East and German West radon study exhibit quite different SDs of radon exposure of about 325, 101, l07 and 47 Bq/m³ among controls, respectively, and an exposure mean of 212, 56, 74 and 49 Bq/m³ (Auvinen et al. 1996, Darby et al. 1998, Wichmann et al. 1999, Kreienbrock et al. 2001). It is therefore of interest, how the bias from

error in the primary predictor depends on the mean and the variance of the predictor distribution.

Since correcting for multiplicative Berkson error involves entering $e^{0.5\sigma^2_{EMB_1}} Z_1$ instead of Z_1, there is no dependence on any exposure distribution parameter and thus no potential for differential bias.

Additive Berkson error induces no bias anyway.

Correcting for additive classical error involves entering $(1-\sigma^2_{EA_1}/\sigma^2_{Z_1})Z_1$ instead of Z_1. The bias therefore depends on the exposure variance, $\sigma^2_{Z_1}$, which is illustrated in *Figure 16*, but not on the exposure mean. It can be seen that there is a great dependence of the bias, indicated by the difference between the naive OR estimate and the 1.12-line, on the SD of the predictor variable, which is of about the same dimension as the dependence on error size.

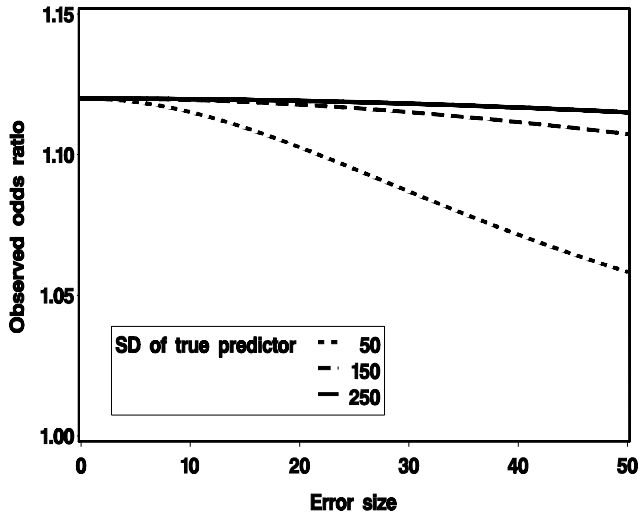

Figure 16: Theoretically observed odds ratio (per 100 units of the predictor variable) for increasing size of additive classical error in the predictor for various standard deviations (SD) of the predictor's distribution with underlying true odds ratio of 1.12 (per 100 units).

The bias from multiplicative classical error cannot be formulated explicitly as for the additive error (see 3.2.1), therefore simulations with varying predictor distribution parameters were necessary (see Heid et al., in press). *Figure 17* displays the GM of the naive OR estimates of 1000 simulated case-control studies and 95% confidence interval for the simulation precision for varying predictor distribution parameters, given a true OR of 1.12. It can be seen that there is no potential for differential bias between groups of varying GM of the predictor distribution (see *Figure 17a*). The bias increases, however, with decreasing GSD of the predictor distribution, which implies according to 2.1.1.1 that the bias increases with decreasing CV, with increasing mean or with decreasing SD of the predictor.

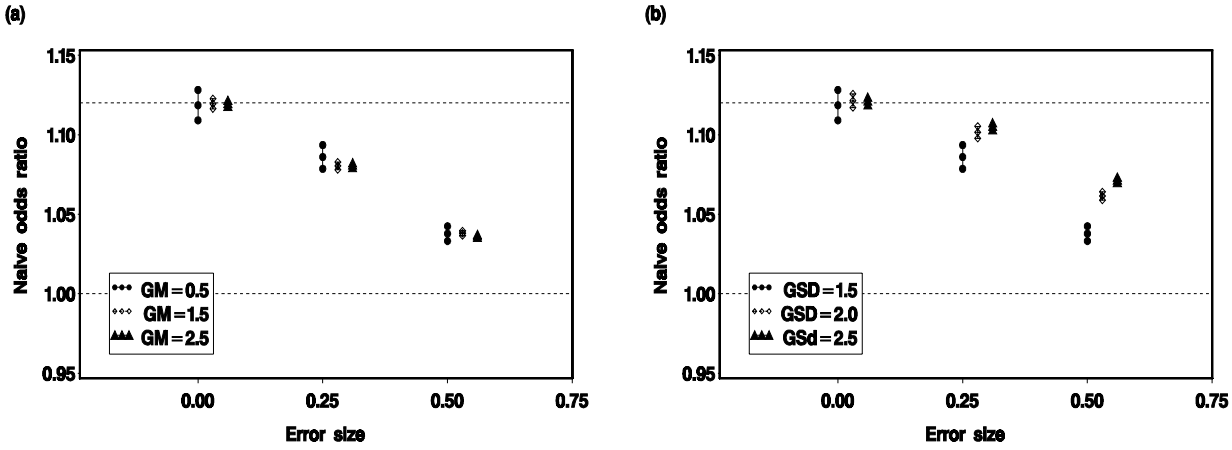

Figure 17: Simulations: Average naive odds ratio estimates (per 100 units of the predictor) and 95% confidence intervals for the simulation precision for increasing size of multiplicative classical error in the predictor with varying (a) geometric mean (GM), (b) geometric standard deviation (GSD) of the true predictor distribution. The underlying true odds ratio is 1.12 (per 100 units).

Inconsistency of naive OR estimates in groups with different predictor variances are therefore possibly explained by differential bias from additive classical error in the predictor. Inconsistency of naive OR estimates in groups with different GSD - or in other words with different CV - of the predictor might be explained by differential bias from multiplicative classical error in the predictor. Given the SD of radon exposure of the European radon studies (see above) and an additive classical error, the bias in the naive OR estimates, which solely depends on the predictor variance, is the largest for the German West study and the smallest for the Finish study. The CVs of the European radon studies mentioned above, the Finish, the

English, the German East and the German West study, are 1.53, 1.80, 1.45 and 0.96 among controls. Thus, for multiplicative classical error, the bias in the naive OR estimates is the largest for the German West study, but the smallest for the English study, since the English study has the largest SD of radon exposure relative to the radon exposure mean. The SD of radon exposure between West and West-High differ (47 versus 66 Bq/m³ among controls); also the CVs differ (0.95 versus 1.11). It is thus theoretically possible, that this difference in the exposure distribution results in differential bias from additive or multiplicative error and explains some of the discrepancy of the naive OR estimates.

3.2.2 Including an error-free covariate

3.2.2.1 Impact on bias from error in the primary variable

Now, the impact of including an error-free covariate into the exposure-disease model on the bias on the effect parameter of primary interest from error in the primary variable is analysed.

If the variable of primary interest is measured with <u>additive classical error</u> and the included covariate is error-free, then γ_{22} equals unity, γ_{21} is zero, and the exposure-disease model accounting for such error is $\text{Logit}(p) = \tilde{\alpha} + \beta_{T_1}\gamma_{11}Z_1 + (\beta_{T_2} + \beta_{T_1}\gamma_{12})Z_2$ (model (13)). The adjustment for the covariate, $(\beta_{T_2} + \beta_{T_1}\gamma_{12})Z_2$ instead of $\beta_{Z_2}Z_2$, is thus altered by the error in the primary variable, which involves, however, only a different scaling of the covariate and this has no impact on the effect estimate of primary interest. However, the bias on β_{Z_1} from error in Z_1 is slightly modified (the reliability factor is now γ_{11} instead of γ_1), if the predictors are correlated. If the predictors are uncorrelated, the model reduces to $\text{Logit}(p) = \tilde{\alpha} + \beta_{T_1}\gamma_1 Z_1 + \beta_{T_2}Z_2$, the error correction is the same as in the univariate case (see 3.2.1), and the bias on β_{Z_1} is not affected by including the uncorrelated covariate.

Since γ_{11} turns unity and γ_{21} is zero for <u>additive Berkson error</u>, there is still no bias from additive Berkson error in the variable of primary interest.

If the variable of primary interest is measured with <u>multiplicative classical error</u>, then m_2 and γ_{22} equal unity, γ_{21} equals zero, and the exposure-disease model correcting for such error is

$\text{Logit}(p) = \alpha + \beta_{T_1} m_1 Z_1^{\gamma_{11}} Z_2^{\gamma_{12}} + \beta_{T_2} Z_2$ (model (12)). The adjustment for Z_2 is thus not influenced by the error in the first variable. However, including the covariate involves replacing Z_1 by $m_1 Z_1^{\gamma_{11}} Z_2^{\gamma_{12}}$, a multiplicative interaction between the predictors, instead of by $m_1 Z_1^{\gamma_1}$. Therefore, inclusion of a covariate, even if error-free, can alter the bias on β_{Z_1}. But if the covariate is uncorrelated with Z_1, the model reduces to $\text{Logit}(p) = \alpha + \beta_{T_1} m_1 Z_1^{\gamma_1} + \beta_{T_2} Z_2$, the error is the same as in the univariate case, and the bias on β_{T_1} remains unmodified by including the uncorrelated covariate.

The model correcting for <u>multiplicative Berkson error</u> in the variable of primary interest including an error-free covariate is $\text{Logit}(p) = \alpha + \beta_{T_1} \cdot e^{0.5\sigma^2_{EMB_1}} Z_1 + \beta_{T_2} Z_2$. Thus, the inclusion of the covariate, if correlated or not, does not influence the bias on β_{T_1}.

3.2.2.2 Dependence of the bias on the correlation

It is now explored, how the bias on the effect parameter of primary interest from error in the primary variable depends on the correlation of an error-free covariate included into the model.

If the primary variable is measured with <u>Berkson error</u>, the correlation between the exposures has no impact on the bias on β_{Z_1}, since the correction for Berkson error does not involve any of the γ_{ij} (i,j=1,2).

If the primary variable is measured with <u>additive classical error</u>, there is at most a small impact of the correlation on the bias given the covariate is error-free: The naive beta-estimate is on average $\gamma_{11} \cdot \beta_{X_1}$, and the difference of γ_{11} (non-zero correlation) and γ_1 (zero correlation) is rather small. The sign of the correlation has definitely no effect, since the correlation appears only as square in γ_{11}. In fact, for parameters reflecting the radon studies' situation, *Figure 15* illustrates that the bias from additive error in the variable of primary interest solely depends on error size and is not influenced by the correlation of the primary predictor with an error-free covariate. The size of the effect of the covariate has no impact

either.

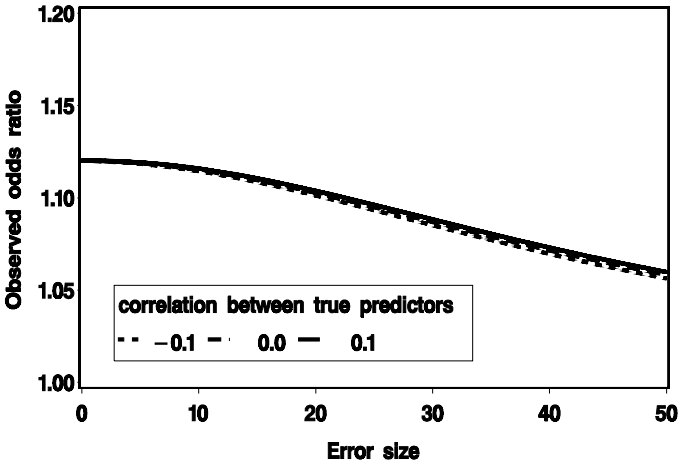

Figure 18: Theoretically observed odds ratio (per 100 units of the primary predictor) for increasing size of additive classical error in the primary predictor for various correlations between the primary predictor and an error-free covariate. Underlying true odds ratio in the primary predictor is 1.12 (per 100 units). The odds ratio of the covariate has no impact here. The SD of the primary variable and the covariate are 50 and 15, respectively.

If the primary variable is measured with multiplicative classical error, the impact of the correlation on the bias on the effect estimate of primary interest cannot be depicted explicitely as for the additive error and is data dependent.

3.2.3 Including an error-prone covariate

3.2.3.1 Impact on the effect estimate of primary interest

Now, the impact of error in the covariate on the effect parameter of primary interest is studied, when the variable of primary interest is error-free.

Including a covariate measured with Berkson error does not induce any bias. The exposure-disease model accounting for additive Berkson error in any predictor does not at all involve the Berkson error size. For multiplicative Berkson error in the covariate, the exposure-disease model, $\text{Logit}(p) = \alpha + \beta_{T_1} Z_1 + \beta_{T_2} e^{0.5\sigma^2_{EMB_2}} Z_2$, indicates that there is no bias induced on the effect parameter of primary interest, β_{T_1}, since scaling of the covariate by a constant, $e^{0.5\sigma^2_{EMB_2}}$, does not affect β_{T_1} in the logistic regression model.

If the covariate is measured with additive classical error, the exposure-disease model accounting for such error, $\text{Logit}(p) = \tilde{\alpha} + (\beta_{T_1} + \beta_{T_2}\gamma_{21})Z_1 + \beta_{T_2}\gamma_{22} Z_2$ (model (13), indicates that the error in the covariate induces bias on the effect estimate of primary interest. The different adjustment for the covariate, $\gamma_{22} Z_2$ instead of Z_2, does not affect β_{T_1}, since scaling of a covariate does not matter.

If the covariate is measured with multiplicative classical error, the exposure-disease model accounting for such error, $\text{Logit}(p) = \alpha + \beta_{T_1} Z_1 + \beta_{T_2} m_2 Z_2^{\gamma_{22}} Z_1^{\gamma_{21}}$ (model (14)), indicates that. the covariate is adjusted for differently, namely by $m_2 Z_2^{\gamma_{22}} Z_1^{\gamma_{21}}$ instead of for Z_2. Since scaling of the adjusting covariate does not influence β_{T_1}, this is equivalent to adjusting for $Z_2^{\gamma_{22}} Z_1^{\gamma_{21}}$, a multiplicative interaction between the exposures. .Therefore, errors in the covariate have the potential to alter the bias on β_{T_1}, if the covariate is correlated. Correcting for errors in an uncorrelated covariate involves adjusting for $Z_2^{\gamma_2}$ instead of Z_2, which impacts the bias on β_{T_1} less.

3.2.3.2 Dependence of the bias on the correlation

It is of interest how the bias on the effect parameter of primary interest from error in the covariate depends on the correlation between the covariate and the primary predictor.

It was already stated that Berkson error in the covariate does not induce any bias on the effect of primary interest. Therefore only classical error is considered in the following.

For homoscedastic additive classical error in the covariate, the naive estimate of the effect of primary interest, β_{Z_1}, is on average $\beta_{X_1} + \gamma_{21}\beta_{X_2}$. Given adverse effects of the predictors, β_{X_1} and β_{X_2} >0, $\beta_{X_1} + \gamma_{21}\beta_{X_2}$ is smaller than the true coefficient for negatively correlated covariates (bias towards the null) and larger for positive correlation (bias away from the null). For zero correlation, there is no bias. The extent of the bias, depends on the size of the covariate's effect, β_{X_2}, the size of the correlation, the size of the error and the size of the exposure variances of both predictor variables (on the log-scale).

Figure 19 illustrates the dependence of the bias on the correlation for small and large disease risk of the confounder. It can be seen that the bias from the error in the covariate is negligible, if the disease risk of the confounder is small (see *Figure 19a*). If the disease risk of the confounder is large (*Figure 19b*), substantial bias in the effect of primary interest is found from error in the covariate, if the correlation is non-zero; there is no bias for zero correlation. As deducted analytically (above), the bias is towards the null for negative correlation and away from the null for positive correlation.

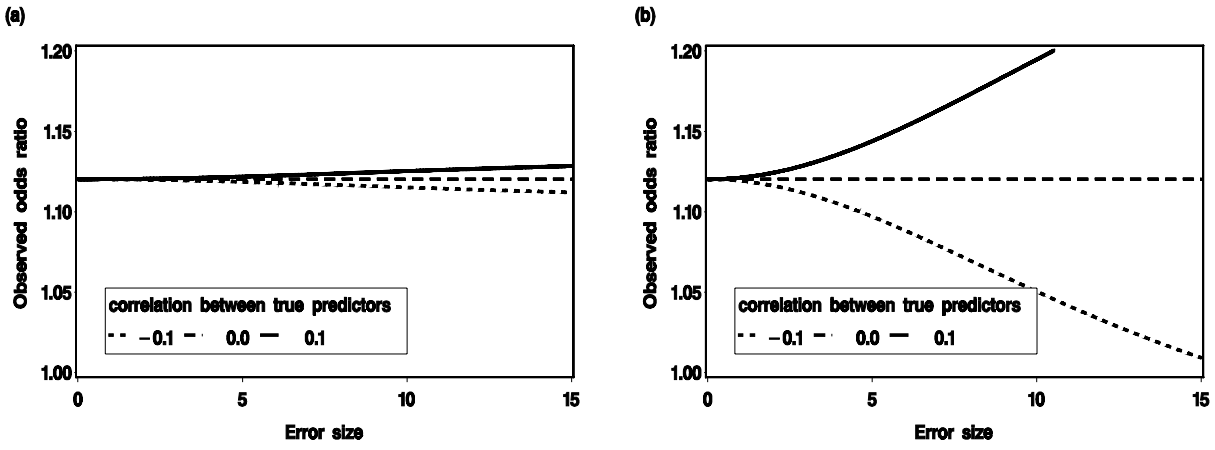

Figure 19: Theoretically observed odds ratio (per 100 units of the primary predictor) for increasing size of additive classical error in the covariate for various correlations between the primary predictor and the covariate. The underlying true odds ratio in the primary predictor is 1.12 (per 100 units) and the true odds ratio in the covariate is (a) 1.05 and (b) 2.0 (per 10 units). The SD of the primary predictor and the covariate are 50 and 15, respectively.

The effect of the correlation on the bias from multiplicative classical type error in the covariate cannot be formulated explicitly as for the additive error, but simulations indicate (not displayed) that the direction of the bias from multiplicative classical error depends in a similar way on the correlation as the bias from additive classical error.

3.2.4 Confidence intervals

The larger the variance of the predictor, $Var[Z_1]$, in the logistic regression model, the larger the power and the more narrow are the confidence intervals. If the variance of the predictor is inflated by error, the confidence intervals are spuriously narrow. The regression calibration method involves replacing the observed predictor, Z_1, by $E[X_1|Z_1]$. The impact of the error correction (assuming the error size is known) on the confidence intervals can thus given by the proportion of $Var[E[X_1|Z_1]]$ to $Var[Z_1]$.

For the additive classical error in the variable of primary interest, Z_1, and assuming no correlation of Z_1 to any covariate, the variance of $E[X_1|Z_1]$ is

$$\gamma_1^2 \cdot Var[Z_1]$$

and thus smaller than $Var[Z_1]$, which implies wider confidence intervals after correction. The variance of $E[X_1|Z_1]$ equals, in fact, the variance of the true predictor, $Var[X_1]$, implying that the corrected confidence intervals are as wide as the confidence intervals would have been, if the true predictor was known and plugged into the exposure-disease model. Note that, when the uncertainty in error size is additionally taken into account, the confidence intervals after correction are wider than the confidence intervals based on the true predictor.

In the case that Z_1, measured with additive classical error, is correlated to a covariate Z_2, the variance of $E[X_1|(Z_1, Z_2)]$ is

$$\gamma_{11}^2 \cdot Var[Z_1] + \gamma_{12}^2 \cdot Var[Z_2]$$

and thus larger than if the predictors were uncorrelated ($\gamma_{12}^2 \cdot Var[Z_2]$ is larger than zero). Including a correlated covariate thus results in more narrow confidence intervals after correction than without the covariate. This is plausible, since the inclusion of a correlated

covariate increases the information on X_1. Note that the error in the covariate does not appear in the variance formulation, which implies no impact of the error in the covariate on the width of the confidence intervals of the effect estimate of primary interest.

For multiplicative classical error in Z_1 and assuming no correlation of Z_1 to any covariate, the variance of $E[X_1|Z_1]$ is $m_1^2 \cdot Var[Z_1^{\gamma_1}]$. Since $Z_1^{\gamma_1} = e^{\gamma_1 \log Z_1}$ is lognormally distributed, $LN(\gamma_1 \cdot \mu_{\log Z_1}, \gamma_1^2 \cdot \sigma^2_{\log Z_1})$, the variance of $E[X_1|Z_1]$ is

$$(e^{2\mu_{\log Z_1}})^{(\gamma_1-1)} \cdot (e^{\sigma^2_{\log Z_1}})^{(\gamma_1^2-1)} \cdot \frac{((e^{\sigma^2_{\log Z_1}})^{\gamma_1^2} - 1)}{(e^{\sigma^2_{\log Z_1}} - 1)} \cdot Var[Z_1].$$

Since γ_1 is smaller than unity, the first three factors are smaller than unity, and the variance of $E[X_1|Z_1]$ is thus smaller than the variance of Z_1. This implies that the confidence intervals are wider after correction.

In the case that Z_1 is measured with multiplicative classical error and correlated to Z_2, the variance of $E[X_1|(Z_1, Z_2)]$ is $m_1^2 \cdot Var[Z_1^{\gamma_{11}} Z_2^{\gamma_{12}}]$, where $Var[Z_1^{\gamma_{11}} Z_2^{\gamma_{12}}]$ is

$$(e^{2\mu_{\log Z_2}})^{\gamma_{12}} \cdot (e^{\sigma^2_{\log Z_2}})^{\gamma_{12}^2} \cdot e^{2\gamma_{11}\gamma_{12}r} \cdot \frac{((e^{\sigma^2_{\log Z_1}})^{\gamma_{11}^2} (e^{\sigma^2_{\log Z_2}})^{\gamma_{12}^2} e^{2\gamma_{11}\gamma_{12}r} - 1)}{((e^{\sigma^2_{\log Z_1}})^{\gamma_{11}^2} - 1)} \cdot Var[Z_1^{\gamma_{11}}].$$

The first factor is smaller [larger] than unity for negative [positive] correlation; the other factors are larger than unity. Thus, the variance of $E[X_1|(Z_1, Z_2)]$ tends to be larger than if the predictors are uncorrelated, but for negative correlations, this depends on the distribution of the covariate. As before, the error in the covariate does not impact the width of the confidence intervals.

For additive Berkson error, the variance of $E[X_1|Z_1]$ is exactly the variance of Z_1, since γ_1 turns unity, which implies that there is no effect on the confidence intervals. For multiplicative Berkson error, the variance of $E[X_1|Z_1]$ is $e^{\sigma^2_{EMB_1}} \cdot Var[Z_1]$, which is larger than $Var[Z_1]$ implying more narrow confidence intervals after correction. Note that, when the uncertainty in the error size is additionally taken into account, the confidence intervals widen, resulting in wider confidence intervals than if the true predictor variable had been available.

The question of statistical significance is therefore not very much influenced by correction for errors in the predictors: The absolute value of the effect estimates corrected for classical type error are usually larger after correction, but the confidence intervals widen, so that the null hypothesis is generally included [excluded] if it was included [excluded] before the correction. On the other hand, the absolute value of the effect estimates corrected for multiplicative Berkson error are usually smaller after correction, but the confidence intervals are more narrow, so that the statistical significance tends to be the same before or after correction. However, borderline significance might be changed by error correction.

It is interesting that the width of the confidence intervals for the effect of primary interest are not affected by errors in the covariate. However, the width of the confidence intervals of the effect of the covariate enlarges with error in the covariate, which implies that the effect of the covariate is less precisely estimated giving room for residual confounding due to greater uncertainty in the controlled for confounding effect.

3.2.5 Approximate likelihood method by Reeves et al.

The exposure-disease models accounting for errors in the predictors derived by the approximate likelihood (AL) method described by Reeves et al. (1998) deviate from models derived by the regression calibration (RC) method only by division of the square root s (see 2.1.4). Exploring the squareroot thus provides comparability of these methods.

Only if both predictor variables are error-prone and the correlation between the predictors is negative, the squareroot can possibly fall short of unity. Otherwise, the squareroot is larger than unity, which implies that $\frac{T_i}{s}$ is smaller than $T_{i\cdot}$, the variance of the predictor after correction with the AL method is smaller than after correction with the RC method, and the effect of primary interest corrected with the AL method is larger than the effect corrected with the RC method.

For <u>additive or multiplicative classical error in the primary variable</u>, it is $\frac{T_i}{s} < T_i < Z_i$. The correction with the AL method is thus larger than the correction with the RC approach,

leading to larger effect estimates.

For additive Berkson error in the primary variable, it is $\frac{T_i}{s} < T_i = Z_i$. The effect estimate corrected with the AL method is larger than the naive estimate. Applying the RC method has no effect for additive Berkson error.

For multiplicative Berkson error in the primary variable, it is $Z_i < \frac{T_i}{s} < T_i$ or $\frac{T_i}{s} < Z_i < T_i$. The correction for multiplicative Berkson error with the AL method may thus be in the same direction as the RC correction, but smaller (leading to effect estimates that are smaller after correction when compared to the naive estimate, but higher when compared to the RC method); or the correction might be in the opposite direction (leading to effect estimates that are higher after correction compared to the naive estimate and thus also higher than the effect estimates corrected by the RC method).

Correcting for classical error in a uncorrelated covariate or correcting for additive Berkson error, which has no effect with the RC method, has a small impact on the effect estimate with the AL method due to the non-unity of the squareroot.

It turns out that s is very close to unity for small beta-coefficients or small errors (thus small $\sigma_{ER_{ij}}{}^2$): Unity is the first order approximation of s by the power series expansion[2] $s = 1 + \frac{1}{2}k^2\beta_1{}^2 m_1{}^2(e^{\sigma_{ER_1}} - 1) \cdot Z_1{}^{2\gamma} - \ldots$ for multiplicative error and $s = 1 + \frac{1}{2}k^2\beta_1{}^2\sigma_{ER_1} - \ldots..$ for additive error in primary predictor (without any correlated covariate included); the second and further terms of the expansion are close to zero if the beta-coefficients and σ_{ER_i} are small.

Generally, s increases - and thus the difference between the AL and RC method increases -

- with increasing size of the measurement error,
- with increasing risk of the error-prone predictor (primary predictor or covariate),

---

[2] Generally, $\sqrt{1+a} = 1 + \frac{1}{2}a - \frac{1}{8}a^2 + \ldots$ $1/(1+a) = 1 + 1/2a - 3/8a^2 - \ldots.$

- with decreasing variance (on the log-scale for multiplicative error) of the error-prone predictor (primary predictor or covariate).

Therefore, the difference between the two methods should only become apparent in the relative lung cancer risk estimate due to radon exposure, a rather small risk, when correcting for very large errors in radon exposure, for errors in radon exposure in a study with small radon exposure variance such as the West study, or for errors in a confounder with large disease risk such as smoking.

3.3 Impact of errors in risk factors in the German Radon Studies

Our considerations from 3.1 suggest overall that a multiplicative model combining classical and Berkson type errors as in (3) is applicable, where Z_1 denotes the *radon concentration in the home inhabited by the study participant at index date*, X_1 denotes the *true alpha dose*, Q_1 denotes an interim variable such as *true residential radon exposure*, EM_1 denotes the product of all classical components of error in the radon variable, the *error in radon exposure,* and EMB_1 denotes the product of all Berkson components of error in the radon variable, the *error in using radon exposure as a surrogate of true alpha dose* (see 3.1.3.1.1). Further, Z_2 denotes the *assessed packyears+1*, X_2 denotes *true lung dose from inhaled carcinogens released by smoking*, Q_2 denotes an interim variable such as *true packyears+1*, EM_2 denotes the product of all classical components of error in the smoking variable, the *error in packyears*, and EMB_2 denotes the product of all Berkson components of error in the smoking variable, the *error from using packyears as a surrogate for true lung dose from inhaled carcinogens released by smoking* (see 3.1.3.1.2).

OR estimates accounting for the error in radon exposure and packyears are derived by applying the regression calibration method (see 2.1.3). Hereby, the Berkson error variance in the error model (3) is zero and the model reduces to a pure multiplicative classical error for both predictors, radon exposure and packyears+1, with the true predictors being true radon exposure and truly smoked packyears+1. Further, it is explored, whether the choice of an additive error model would make a great difference. Additionally, the OR estimate due to alpha dose (again, adjusted for the smoked packyears) are of interest and thus a Berkson error for the error in using radon exposure as a surrogate for alpha dose needs to be included. It

is also of interest whether including a Berkson error in the smoking variable would change the results, in order to adjust for the lung dose of inhaled carcinogens rather than for packyears. Finally, all these analyses are repeated applying the approximate likelihood approach described in 2.1.4.

3.3.1 Correcting for errors in radon exposure and packyears (multiplicative classical error)

3.3.1.1 Correcting the correlation for multiplicative classical error

Since the correlation between the two predictors, radon and smoking, plays an important role in the confounding potential of smoking and in the correction of the risk estimates (see 3.2.3.2), the correlations (on the log-scale) corrected for various sizes of multiplicative classical error according to equation (6) are presented in *Table 13*.

Table 10: German radon studies: Pearson correlation coefficient between log of radon exposure and log(packyears+1) among controls corrected for multiplicative classical errors of various sizes. Error size is given as SD of log of error and in parentheses as percentage of the predictor variance observed among controls explained by the error (with the West study as reference for radon exposure and the East study for packyears).

error in radon	*error in packyears*	*West*	*West High*	*East*
0 (0%)	0	-0.060	0	-0.060
0.3 (25%)	0	-0.070	0	-0.068
0.4 (50%)	0	-0.081	0	-0.077
0	0.5 (12%)	-0.063	0	-0.064
0	0.7 (25%)	-0.067	0	-0.069
0	1.0 (50%)	-0.079	0	-0.084
0.3 (25%)	0.5 (12%)	-0.073	0	-0.073
0.4 (50%)	0.5 (12%)	-0.086	0	-0.082

It can be seen that the „true“ correlation increases with increasing error size and is quite

similar for West and East. Whereas the observed correlation is the same in the West and the East study, the correlation corrected for error in radon exposure is a little larger in West than in East due to the smaller variance of observed radon exposure (on the log scale). The correlation corrected for error in packyears is slightly larger in the East study due to the smaller the variance of observed packyears (on the log-scale).

3.3.1.2 Correcting the OR estimates for multiplicative classical error: Full data

Figure 20 displays the derived OR estimates for 100 Bq/m³ increase in radon exposure for various sizes of multiplicative classical error in radon exposure (assuming no error in packyears), for various sizes of multiplicative classical error in packyears (assuming no error in radon exposure), and for various sizes of multiplicative classical error in radon exposure assuming a fixed size of 0.5 (SD of the log of the error) for the error in packyears. The regression calibration method accounting for the correlation between radon exposure and packyears is hereby applied.

For zero error (*Figure 20a,b*), the „naive" ORs of 0.97, 1.12, and 1.04 can be seen for the West, the West-High, and the East study, respectively (as in *Figure 20*).

Accounting for errors of non-zero size in radon exposure assuming that packyears are error-free (*Figure 20a)*, overall higher OR estimates are found. The ORs accounting, for example, for an error of size 0.4 in radon exposure, that is an error explaining about 50% [40%] of the observed variance of radon exposure (on the log-scale) in West [East], are 1.02, 2.09 and 1.11 for West, West-High and East, respectively. In fact, the OR in the West study exceeds unity for errors in radon exposure larger than 0.3, even though the uncorrected estimate was 0.97.

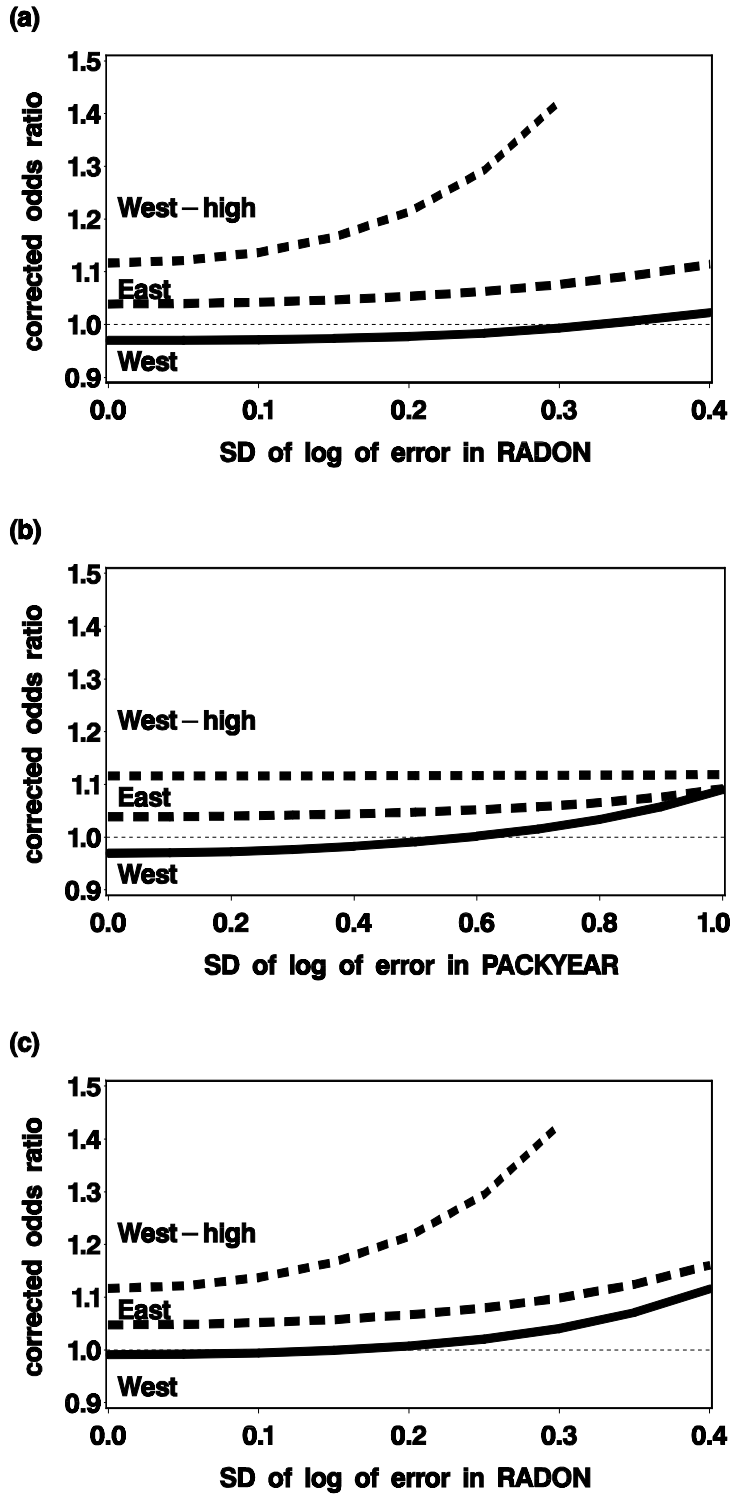

Figure 20: German radon studies: Odds ratio estimates corrected for various sizes of multiplicative classical error (a) in radon exposure, (b) in packyears, and (c) in radon exposure and a fixed error in packyears of size 0.5 with the correlation between radon exposure and packyears taken into account.

Assuming that radon exposure is error-free and accounting for moderate errors in packyears of size 0 to 0.5 (*Figure 20b)*, the OR of West High shows no change, the OR of East almost none, and the OR in the West study is slightly larger than the naive OR. For large errors in packyears of size 0.5 to 1.0, the OR of West High is still without change, which is due to the lack of correlation between radon exposure and packyears, the OR of East increases slightly, and the OR of West increases more rapidly exceeding unity for errors larger than 0.5. The West OR even coincides with the East OR and approaches the West-High OR for an error of 1.0.

*Figure 20*c displays the OR estimates accounting for various sizes of error in radon exposure and a fixed error of 0.5 for the error in packyears, that is an error explaining 12% [10%] of the observed variance of packyears+1 (on the log-scale) in the East [West] study. This provides a generally similar picture as in *Figure 20a*; for West High it is identical. However, the OR estimate for the West study increases more rapidly, now exceeding unity for errors in radon exposure larger than 0.1.

Figure 21 is analogous to *Figure 20a,* but ignoring the correlation between radon exposure and packyears (applying the model (10)). It can be seen that the corrections would then be less conservative, particularly in the West study. Ignoring the correlation in the correction for error in radon exposure of size 0.4 would increase the OR from 1.02 to 1.10 and from 1.11 to 1.13 for West and East, respectively. The ORs of West High are the same as in *Figure 20a* due to the lack of correlation between radon exposure and packyears.

There is basically no effect of correcting for errors in packyears if the correlation is ignored or if none is present (as in West-High); the graphs analogous to *Figure 20b* or *Figure 20c* (not displayed) would only show horizontal lines or the same as *Figure 20c,* respectively.

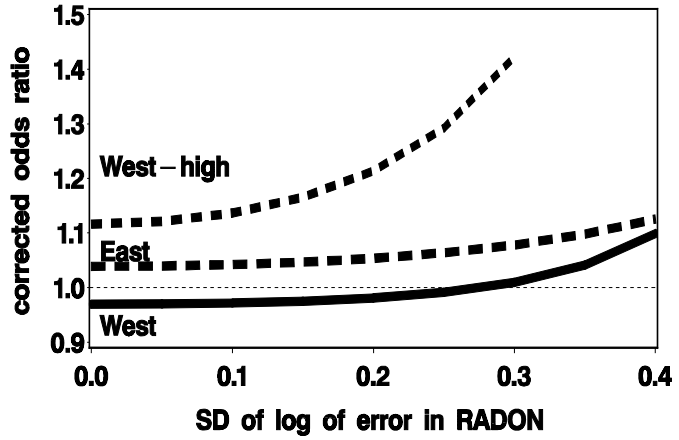

Figure 21: German radon studies: Analogous to Figure 20a, correcting for multiplicative classical type error in radon exposure, with the correlation between radon exposure and packyears ignored.

Table 11 displays OR estimates and 95% confidence intervals reporting the results of the error correction (I) without and (II) with accounting for the correlation between radon exposure and packyears. Comparing column I and II, it can be seen that accounting for the correlation leads to smaller OR when correcting for error in radon exposure (as visible comparing *Figure 20*a and *Figure 21*), but to larger ORs when correcting for error in packyears, which is due to the fact that the correlation is negative. The 95% confidence intervals widen with the correction for increasing error in radon exposure or packyears (compare with 3.2.4). However, for West-High, the lower confidence limit exceeds unity - indicating that the OR estimate is significantly larger than unity on a 5% level - after accounting for errors in radon exposure of size 0.4.

Table 11: German radon studies: Odds ratio estimates for 100 Bq/m³ increase in radon exposure and 95% confidence intervals corrected for multiplicative classical errors in radon exposure and packyears (I) without and (II) with accounting for the correlation between radon exposure and packyears. Error size is given as standard deviation of the log of the error.

error in radon	*error in packyears*	*West*		*West High*	*East*	
		I	*II*	*I*	*I*	*II*
0	0	0.97 (0.81, 1.15)	0.97 (0.81, 1.16)	1.12 (0.87, 1.43)	1.04 (0.96, 1.12)	1.04 (0.96, 1.12)
0.3	0	1.01 (0.74, 1.37)	0.99 (0.73, 1.35)	1.42 (0.92, 2.21)	1.08 (0.93, 1.24)	1.08 (0.93, 1.24)
0.4	0	1.10 (0.67, 1.79)	1.02 (0.62, 1.68)	2.09 (1.05, 4.17)	1.13 (0.89, 1.43)	1.11 (0.88, 1.42)
0	0.5	0.97 (0.82, 1.16)	0.99 (0.83, 1.18)	1.12 (0.87, 1.44)	1.04 (0.96, 1.12)	1.05 (0.97, 1.13)
0	0.7	0.97 (0.82, 1.16)	1.02 (0.85, 1.21)	1.12 (0.87, 1.44)	1.04 (0.97, 1.12)	1.06 (0.98, 1.14)
0	1.0	0.98 (0.82, 1.17)	1.09 (0.91, 1.30)	1.12 (0.87, 1.45)	1.04 (0.97, 1.12)	1.09 (1.01, 1.18)
0.3	0.5	1.02 (0.75, 1.38)	1.04 (0.76, 1.42)	1.43 (0.92, 2.22)	1.08 (0.94, 1.25)	1.10 (0.95, 1.27)
0.4	0.5	1.11 (0.68, 1.82)	1.12 (0.68, 1.83)	2.11 (1.05, 4.23)	1.13 (0.89, 1.44)	1.16 (0.91, 1.48)

3.3.1.3 Correcting for multiplicative classical error: Excluding one outlier

Figure 22 shows the analogy to *Figure 20* with exclusion of the outlier from the West data, which is also part of West-High. The OR estimates for East are repeated unchanged. For zero error, the „naive“ OR estimates of 1.01 and 1.29 (instead of 0.97 and 1.12 as for the full data) can be seen for the West and the West-High study, respectively (*Figure 22*,*a and b*).

The OR estimates corrected for non-zero errors for West and West High are higher when the outier is excluded than with the full data. As expected, the outlying control has attenuated the OR estimate; more remarkably, it has attenuated the OR estimate to even such an extent that the sign of the effect is reversed in the West study (A look at the confidence intervals, however, shows that this reversal is statistically not of significance.). With exclusion of the outlier, the difference of OR estimates between West and East vanish. For an error in packyears larger than 0.6 or for an error in packyears of size 0.5 together with an error in radon of at least 0.2, the West OR even slightly exceeds the East OR (*Figure 22*,*b and c*). A part of the difference between West and West High disappears when correcting for large errors in packyears. Overall, accounting for errors in exposures increases the OR estimates in *Figure 22* in a similar manner as in *Figure 20*.

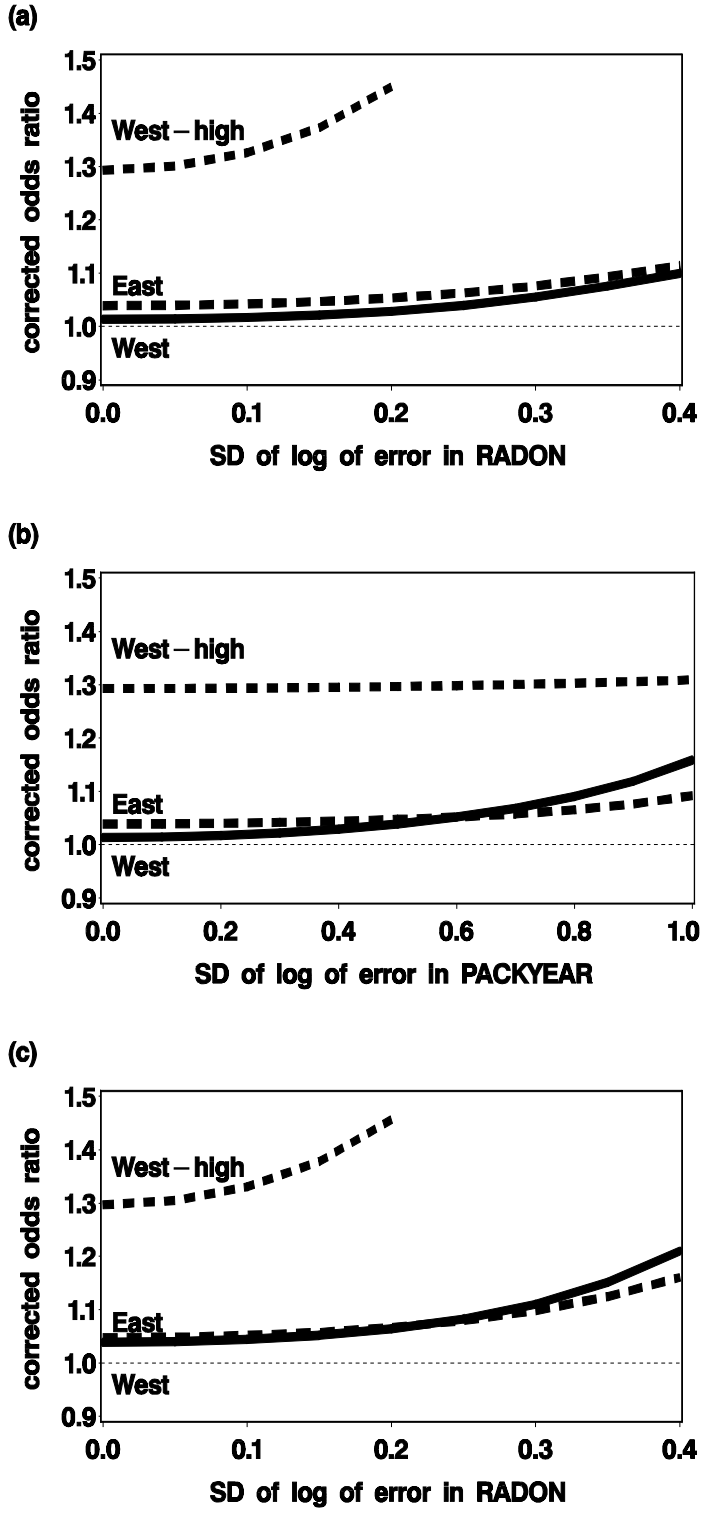

Figure 22: German radon studies: Analogous to Figure 20, correcting for multiplicative classical error accounting for the correlation, without the outlier.

Table 12 shows the analogy to *Table 11* without the outlier. Now the OR estimate for West High becomes significantly higher than unity correcting for an error of 0.3 or larger. The effect of ignoring the correlation between exposures is similar as for the full data.

Table 12: German radon studies: As in Table 11, correcting for multiplicative classical error (I) without and (II) with accounting for the correlation, excluding the outlying control from West (and West High).

error in radon	*error in packyears*	*West*		*West High*
		I	*II*	*I*
0	0	1.01 (0.84, 1.22)	1.01 (0.84, 1.22)	1.29 (0.97, 1.72)
0.3	0	1.08 (0.78, 1.48)	1.05 (0.76, 1.45)	1.78 (1.10, 2.88)
0.4	0	1.19 (0.72, 1.98)	1.10 (0.66, 1.84)	2.86 (1.36, 6.11)
0	0.5	1.02 (0.84, 1.23)	1.04 (0.86, 1.25)	1.30 (0.97, 1.74)
0	0.7	1.02 (0.84, 1.23)	1.07 (0.89, 1.29)	1.30 (0.97, 1.74)
0	1.0	1.03 (0.85, 1.24)	1.16 (0.96, 1.40)	1.31 (0.97, 1.77)
0.3	0.5	1.08 (0.79, 1.49)	1.11 (0.80, 1.53)	1.79 (1.10, 2.91)
0.4	0.5	1.21 (0.73, 2.02)	1.21 (0.72, 2.02)	2.93 (1.37, 6.24)

3.3.2 Sensitivity on the choice of the error model (additive classical error)

3.3.2.1 Correcting for additive classical error: Full data

Figure 23 shows the analogy to *Figure 20* correcting for additive classical error in order to evaluate the sensitivity of the results on the choice of the error model. Hereby, the correlation between radon exposure and packyears is accounted for by using the Spearman correlation coefficient (see 2.1.5).

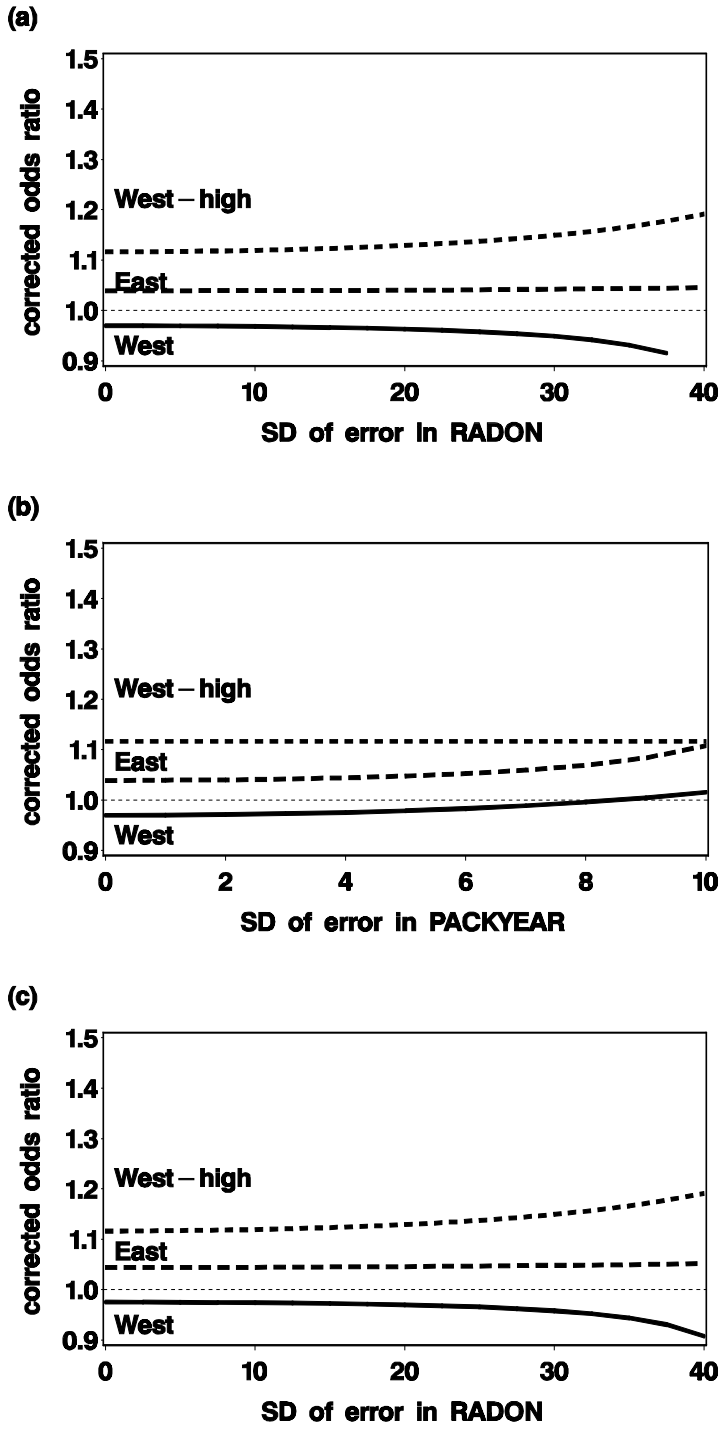

Figure 23: German radon studies: Odds ratio estimates corrected for various sizes of additive classical error in (a) radon exposure, (b) packyears, and (c) radon exposure with fixed size of error in packyears (SD=4) accounting for the correlation (Spearman).

Figure 23a shows that, after correction, the OR estimates in West and West-High are slightly increased („protective“ effect in West, adverse effect in West-High). For example, the OR corrected for an error size of 30, that is an error explaining about 50% of the observed radon exposure variance, is 0.95 for West, 1.19 for West-High. The effect estimates in East remain unchanged, which is due to the large exposure variance in the East study (see 3.2.1.2). These findings are in accordance with the theory that classical error in the predictor induces bias towards the null hypotheses, that is the effect estimates are larger after correction (see 3.2.1). Overall, the impact of correcting for additive error in radon exposure is found to be much smaller than the impact of correcting for multiplicative error.

In *Figure 23*b, it can be seen that correction for additive error in packyears, has no effect on the OR in West-High, which is due to the lack of correlation between the risk factors. It is further found that the OR in both East and West increase with correcting for increasing error size. For West, the OR exceeds unity for large errors (SD of error > 8, an error explaining about 35% of the exposure variance among controls) in packyears.

Correction for an increasing size of additive error in radon exposure and a fixed size for additive error in packyears of size 4 (*Figure 23c*), an error explaining about 10% of the packyear variance among controls, provides a similar picture as *Figure 23a*. The impact of the additive error in packyear of that size is negligible.

Overall, choosing an additive versus a multiplicative error model would not only quantitatively effect the corrected OR estimates for the West, West-High and East study, but also qualitatively in the West study (The effect reversal in the West study is not found for additive error.).

Table 13 shows the OR estimates and 95% confidence intervals correcting for additive classical error (I) without and (II) with accounting for the correlation.

Table 13: German radon studies: OR estimates correcting for additive classical error in the risk factors (I) without and (II) with accounting for the (Spearman) correlation. Error size is given as standard deviation of the error and in parentheses as percentage of exposure variance among controls (with the West study as reference for radon exposure and the East study for packyears) explained by error.

error in radon	*error in packyears*	*West*		*West High*	*East*	
		I	*II*	*I*	*I*	*II*
0	0	0.97 (0.81, 1.15)	0.97 (0.81, 1.16)	1.12 (0.87, 1.43)	1.04 (0.96, 1.12)	1.04 (0.96, 1.12)
20 (25%)	0	0.96 (0.78, 1.19)	0.96 (0.78, 1.19)	1.13 (0.86, 1.59)	1.04 (0.96, 1.12)	1.04 (0.96, 1.12)
30 (50%)	0	0.95 (0.70, 1.28)	0.95 (0.70, 1.28)	1.15 (0.84, 1.58)	1.04 (0.96, 1.13)	1.04 (0.96, 1.13)
0	4 (10%)	0.97 (0.81, 1.16)	0.98 (0.82, 1.16)	1.12 (0.87, 1.43)	1.04 (0.96, 1.12)	1.04 (0.97, 1.13)
0	7 (25%)	0.97 (0.81, 1.16)	0.99 (0.83, 1.18)	1.12 (0.87, 1.43)	1.04 (0.96, 1.12)	1.06 (0.98, 1.14)
0	10 (50%)	0.97 (0.81, 1.16)	1.02 (0.85, 1.21)	1.12 (0.87, 1.43)	1.04 (0.96, 1.12)	1.11 (1.03, 1.19)
20 (25%)	4 (10%)	0.96 (0.78, 1.19)	0.97 (0.78, 1.20)	1.13 (0.86, 1.49)	1.04 (0.96, 1.12)	1.05 (0.97, 1.13)
30 (50%)	4 (10%)	0.95 (0.70, 1.28)	0.96 (0.71, 1.29)	1.15 (0.84, 1.58)	1.04 (0.96, 1.13)	1.05 (0.97, 1.14)

The table entries illustrate that ignoring the correlation in correcting for errors in radon exposure has no impact on the results for additive error. However, ignoring the correlation in correcting for errors in packyears removes all the effect of the correction in West and East; in West-High the correction for error in packyears has no effect due to the lack of the correlation in the first place. The graphs accounting for additive classical error in radon exposure ignoring the correlation (not displayed) would thus show the same picture as *Figure 23a*, since the correlation has no influence on the correction for additive error (in contrast to multiplicative error). The graph analogous to *Figure 23b* but ignoring the correlation (not displayed) would only show horizontal lines, since there is no bias from errors in the confounding variable if the „confounding" variable is uncorrelated.

3.3.2.2 Correcting for additive classical error: Excluding one outlier

Figure 24 displays the analogy to *Figure 23* with exclusion of the outlier in the West study, which is also part of West-High. Since the outlier is not a subject from the East study, the East OR is not affected, but displayed for comparison. Without the outlier, the naive OR estimate of the West study is already larger than unity (1.02), and the correction for additive error in radon exposure increases - as expected from theory- the OR estimates. A comparable increase in the OR estimate can be seen in West-High. The ORs corrected for error in packyears are similar as with the outlier, but on a different level for West and West-High. Thus, the choice of additive error model does not provide a qualitatively different picture of the impact of error correction when compared to a multiplicative error model, if the outlier is excluded.

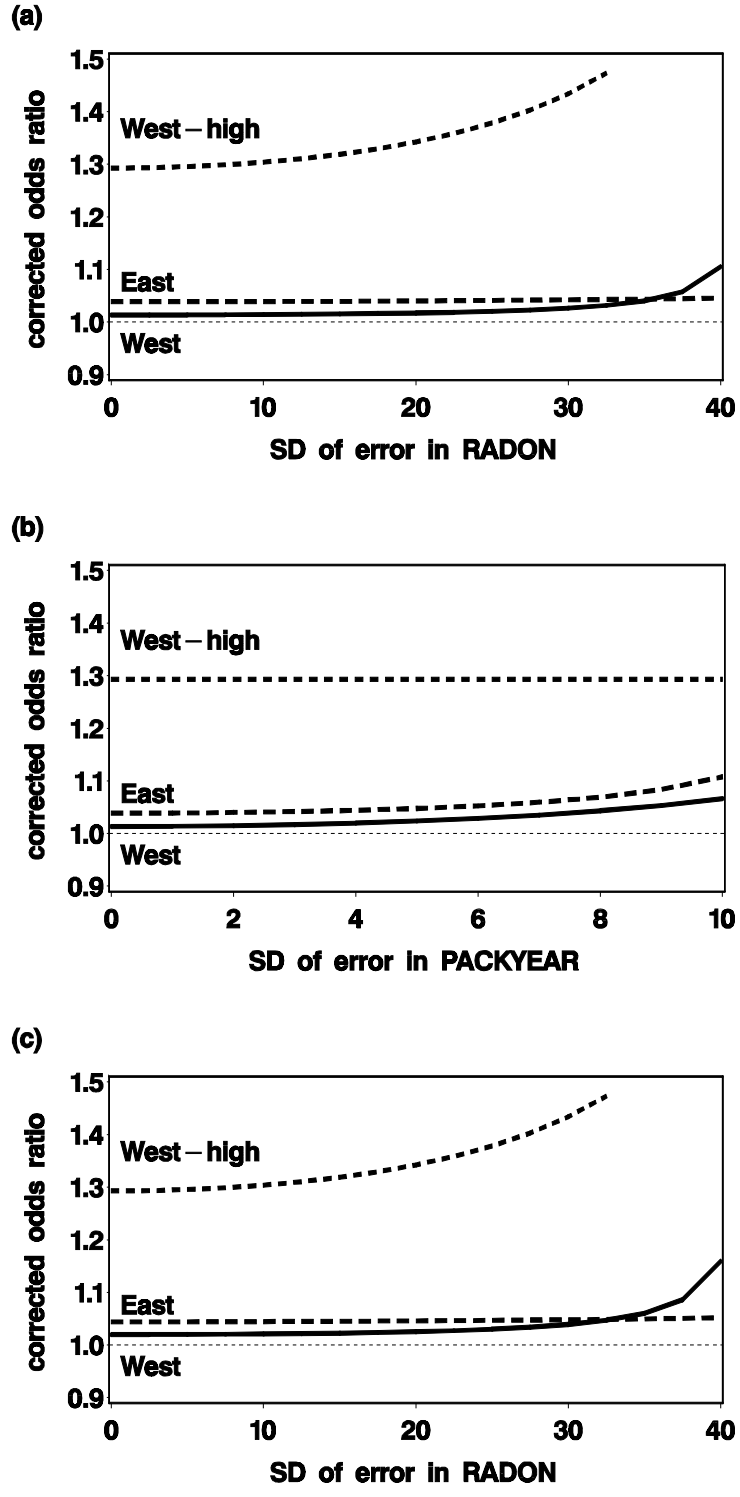

Figure 24: German radon studies: As Figure 23 (additive classical error) excluding the outlier.

3.3.3 Accounting for the use of external exposure as a surrogate for lung dose (Berkson error)

The error made by using radon exposure instead of the truly disease-causing agent, the alpha dose in the lung, is modelled by a Berkson error (see 3.1.3.1.1). Regarding the adjustment for a confounder, it might also be more desirable to adjust for the lung dose of inhaled carcinogens from smoking, instead of adjusting for packyears.

The exposure-disease model accounting for additive Berkson error in the predictor of primary interest or in the confounder has shown that such error does not have any impact on the effect parameter of primary interest (see 3.2.1). Further, multiplicative Berkson error in the covariate does not have any impact on the effect parameter of primary interest either (see 3.2.3). Thus, using packyears instead of the lung dose of inhaled carcinogens from smoking does not induce bias on the estimate of relative lung cancer risk from radon.

3.3.3.1 Correcting for multiplicative Berkson error in radon exposure: Full data

This leaves the impact of multiplicative Berkson error in radon exposure of interest to us, the correction for which is displayed in *Figure 25* without and with additional classical error of fixed size of 0.4. In the correction for Berkson error, the correlation is of no concern (see 3.2.1.2); in the correction for the classical error, the correlation is taken into account.

Figure 25a shows practically no impact of correcting for a 50% Berkson type error (SD on the log-scale of 0.4). With such a Berkson error, the variance of the true lung dose (on the log-scale) would be 1.5 times as large as the observed radon exposure variance. For a Berkson error of 100% (SD on the log-scale of 0.6), that is an error such that the variance of the true lung dose doubles the observed radon exposure variance (on the log-scale), there is no effect on the OR of the West study, the OR estimates in the East and the West-High study are slightly decreased.

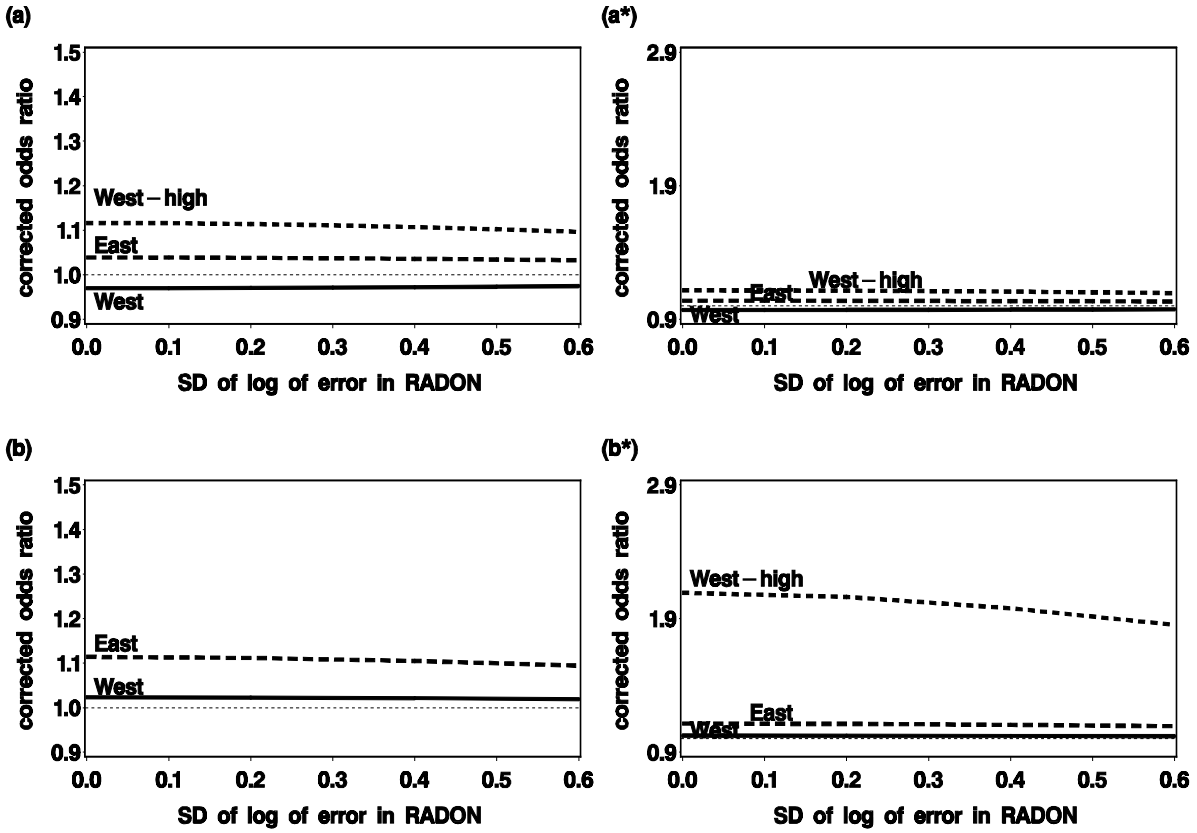

Figure 25: German radon studies: Odds ratio estimates correcting for multiplicative Berkson error in radon exposure (a) without and (b) with additionally correcting for a fixed size of multiplicative classical error (SD on log-scale of 0.4) accounting for the correlation. Since the West-High estimates do not fit onto the graph in (b), the graphs in (a) and (b*) show the same as (a) and (b) with a different scale of the vertical axis.*

Figure 25b shows the impact if the Berkson error is taken into account together with a classical error of 0.4 for the West and the East study; to display the OR estimates of the West-High study, the scale of the y-axis is enlarged (*Figure 25b**). It can be seen that there is no effect other than the effect of the classical error in the West study and that correcting for the Berkson error has a slight OR decreasing impact in the East study and the West-High study, which is, however, negligible when compared to the impact of the classical error.

Table 14 shows the OR estimates visualised in *Figure 25* together with the 95% confidence intervals. It can be seen that the CIs for the estimates corrected for Berkson error are slightly smaller than the naive CIs, which is in accordance with 3.2.4.

Table 14: German radon studies: OR estimates and 95% confidence intervals correcting for multiplicative Berkson error in radon exposure without and with accounting for a fixed multiplicative classical error (SD on log-scale of 0.4). Error size is given as SD of the error and in parentheses as the proportion of the error variance compared to the variance of observed radon exposure (on the log-scale) among controls in the West study.

classical error	*Berkson error*	*West*	*West High*	*East*
0	0	0.97 (0.81, 1.16)	1.12 (0.87, 1.43)	1.04 (0.96, 1.12)
0	0.4 (50%)	0.97 (0.81, 1.16)	1.11 (0.88, 1.39)	1.04 (0.96, 1.11)
	0.6 (100%)	0.97 (0.84, 1.13)	1.10 (0.89, 1.35)	1.03 (0.97, 1.10)
0.4 (50%)	0	1.02 (0.62, 1.68)	2.09 (1.05, 4.17)	1.11 (0.88, 1.42)
	0.4 (50%)	1.02 (0.65, 1.61)	1.98 (1.05, 3.74)	1.10 (0.88, 1.38)
	0.6 (100%)	1.02 (0.67, 1.54)	1.71 (1.04, 2.82)	1.08 (0.91, 1.29)

Since both radon exposure and packyears cannot be assessed without error, the most realistic scenario is a multiplicative classical error in radon exposure of 0.4, a multiplicative classical error in packyears and a multiplicative Berkson error from the using radon exposure as a surrogate for the alpha dose of size 0.6. Correcting for theses errors yields the OR estimates 1.10, 1.87, and 1.13 for the West, the West-High, and the East study, respectively.

3.3.3.2 Correcting for multiplicative Berkson error in radon exposure: Excluding one outlier

The analogous graphs with exclusion of the outlier in the West (and West-High) study is shown in *Figure 26* (the ORs for East are unchanged, but displayed for comparison). *Table 15* shows the OR estimates together with the 95% confidence intervals. Correcting for an error in radon exposure of size 0.4, for an error in packyears of 0.5, and for a multiplicative Berkson error in radon exposure of size 0.6 yields the OR estimates 1.17, 2.45, and 1.13 for the West, the West-High, and the East study, respectively.

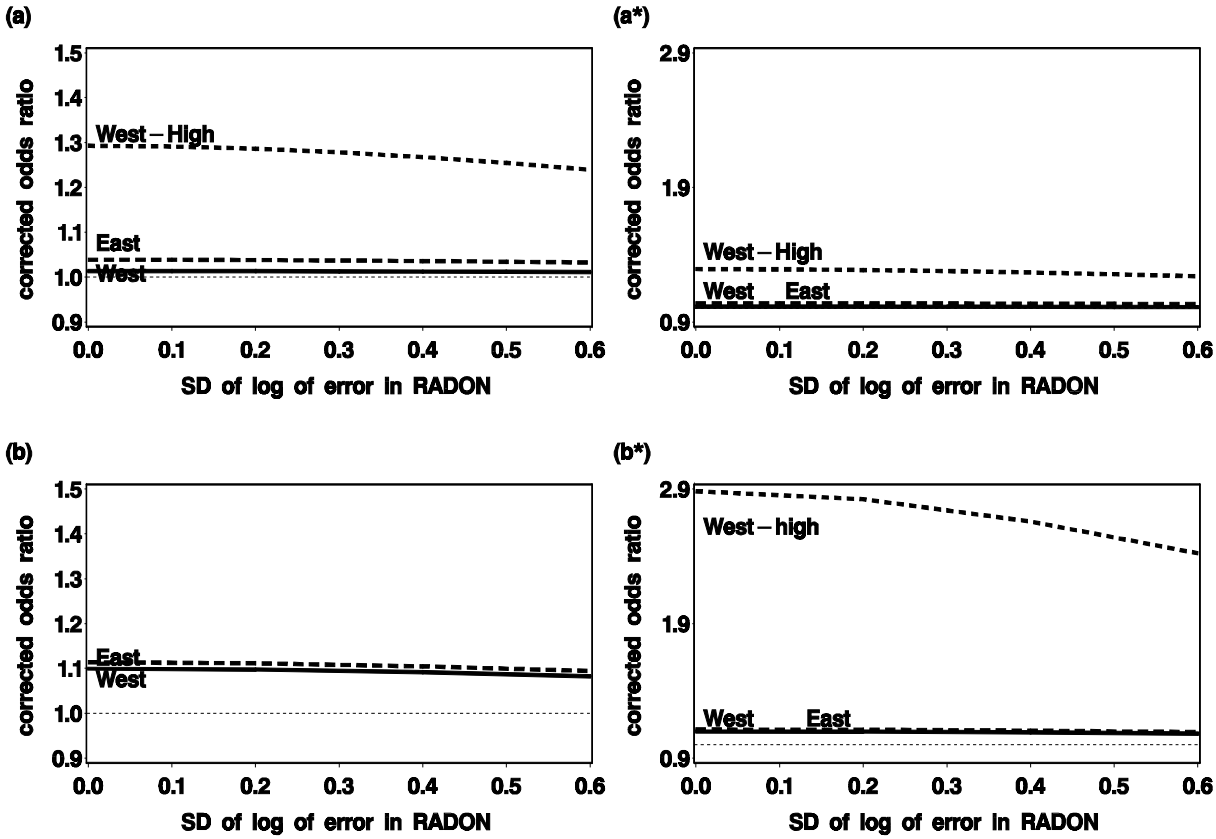

Figure 26: German radon studies: As Figure 25 (correcting for multiplicative Berkson error in radon exposure) without the outlier.

Table 15: German radon studies: As Table 14 (correcting for multiplicative Berkson error in radon exposure) excluding the outlier.

Berkson error	*classical error*	*West*	*West High*
0	0	1.01 (0.84, 1.22)	1.29 (0.97, 1.73)
0	0.4 (50%)	1.01 (0.85, 1.20)	1.27 (0.97, 1.65)
	0.6 (100%)	1.01 (0.84, 1.13)	1.24 (0.97, 1.58)
0.4 (50%)	0	1.10 (0.66, 1.84)	2.88 (1.36, 6.11)
	0.4 (50%)	1.09 (0.68, 1.75)	2.82 (1.35, 5.32)
	0.6 (100%)	1.08 (0.71, 1.66)	2.42 (1.29, 4.54)

3.3.4 Comparison to correction with the approximate likelihood method

3.3.4.1 Correcting for multiplicative classical error

Corrections for multiplicative classical error applying the approximate likelihood (AL) method by Reeves et al. (1998) as described in 2.1.4.2 together with the analogous corrections applying the regression calibration (RC) method as above, are displayed together in *Figure 27*.

It can be seen that, whereas there is basically no difference between the two methods when correcting for errors in radon exposure (see *Figure 27a*), the difference is notable in the West and the West-High study when correcting for errors in packyears: The AL method leads to higher ORs in West-High even for moderate errors in packyears with increasing difference between the methods with increasing error size; for the West study, the AL method leads to slightly higher ORs for errors from about 0.3 to 0.8, but the difference decreases for larger errors, disappears for an error size of 0.9 and reverses (that is the RC method leads to slightly higher ORs than the AL method) of errors larger than 0.9 (see *Figure 27b*).

Correcting for increasing size of error in radon exposure with a fixed error of 0.5 in packyears (see *Figure 27c*), the AL method leads to notably higher ORs for large errors in radon exposure in the West and West-High study. Note that the difference in West would disappear if the fixed error in packyears would be around 0.9 (which is a very large error, see 3.1.4).

For the East study, practically no difference between the two methods is found, except a slightly larger OR with the AL method for large errors in radon exposure and a fixed error of 0.5 in packyears.

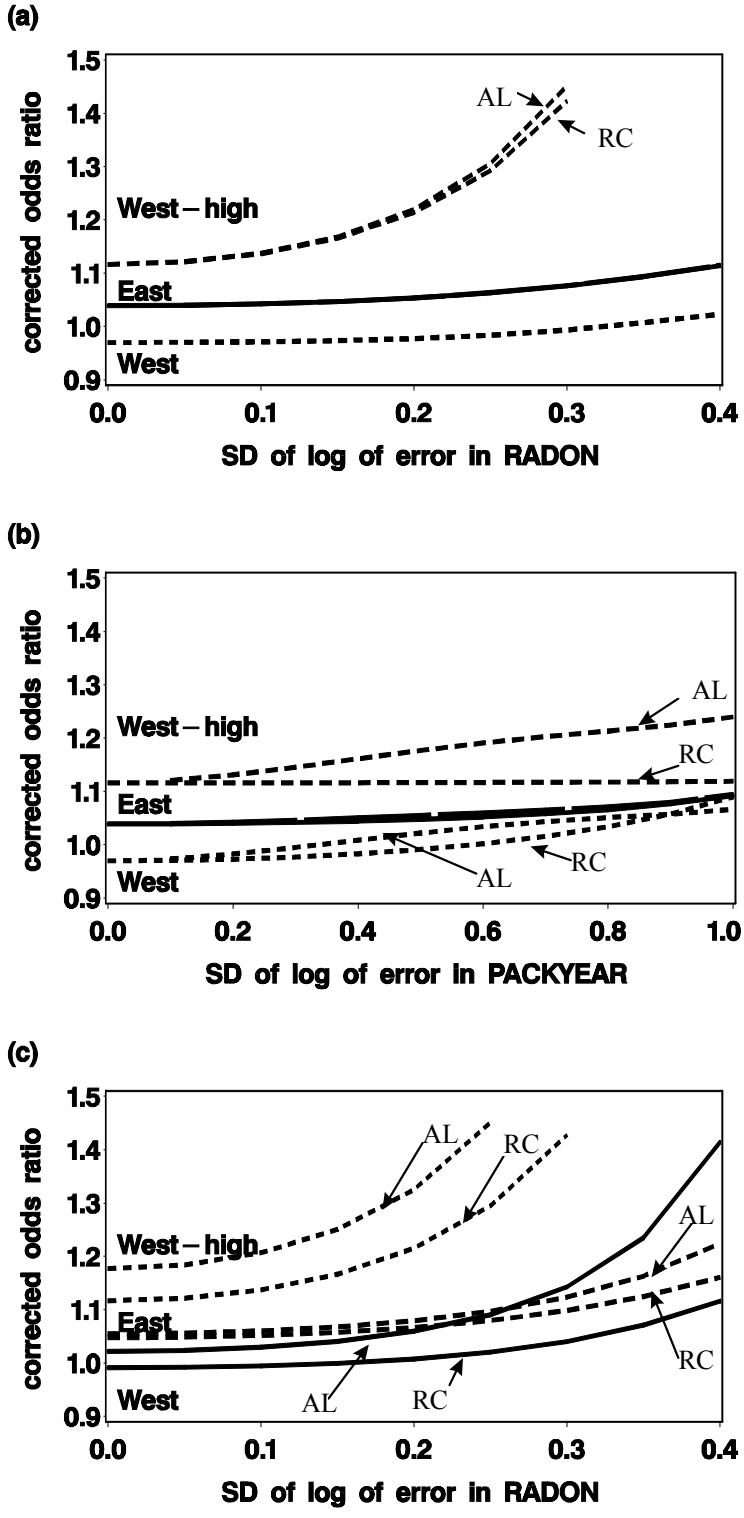

Figure 27: Analogous to Figure 20 (correcting for various sizes of multiplicative classical error in (a) radon exposure, (b) packyears, (c) radon exposure with fixed error in packyear of size 0.5) applying the approximate likelihood (AL) approach versus regression calibration (RC).

3.3.4.2 Correcting for multiplicative Berkson error

Corrections for multiplicative Berkson error applying the approximate likelihood (AL) method as described in 2.1.4.2 together with the analogous corrections applying the regression calibration (RC) method, are displayed together in *Figure 28*.

Assuming no classical error, the correction for multiplicative Berkson error by the two methods shows practically no difference (*Figure 28 a*) for the West and the East study. For the West-High study, the AL method results in slightly higher OR estimates: The correction by the RC method slightly decreases, the AL method slightly increases the OR estimates (for a 100% error of size 0.6, the RC-corrected estimate is 1.10 versus the AL-corrected estimate of 1.12). This difference is explained by the fact that the squareroot s in the AL-model (see 2.1.4.1) is larger than $e^{0.5\sigma^2_{EMB1}}$, which leads to the opposite effect of the correction by the AL method when compared to the RC method (see 3.2.5).

Correcting for multiplicative Berkson error together with a fixed multiplicative classical error of 0.4, the corrections by the two methods do, again, not differ for the West and the East study. However, the difference between the two methods for West-High becomes more pronounced. This is due to fact that the OR estimates are higher in West-High after correction for the classical error (The two methods differ the more the larger the effect of the error-prone variable, as noted in 3.2.5.) and that the OR estimates „at baseline“ differ already for the two methods (The corrections for the classical error result in OR estimates of 2.09 if RC-corrected and 2.24 if AL-corrected, see 3.3.4.1).

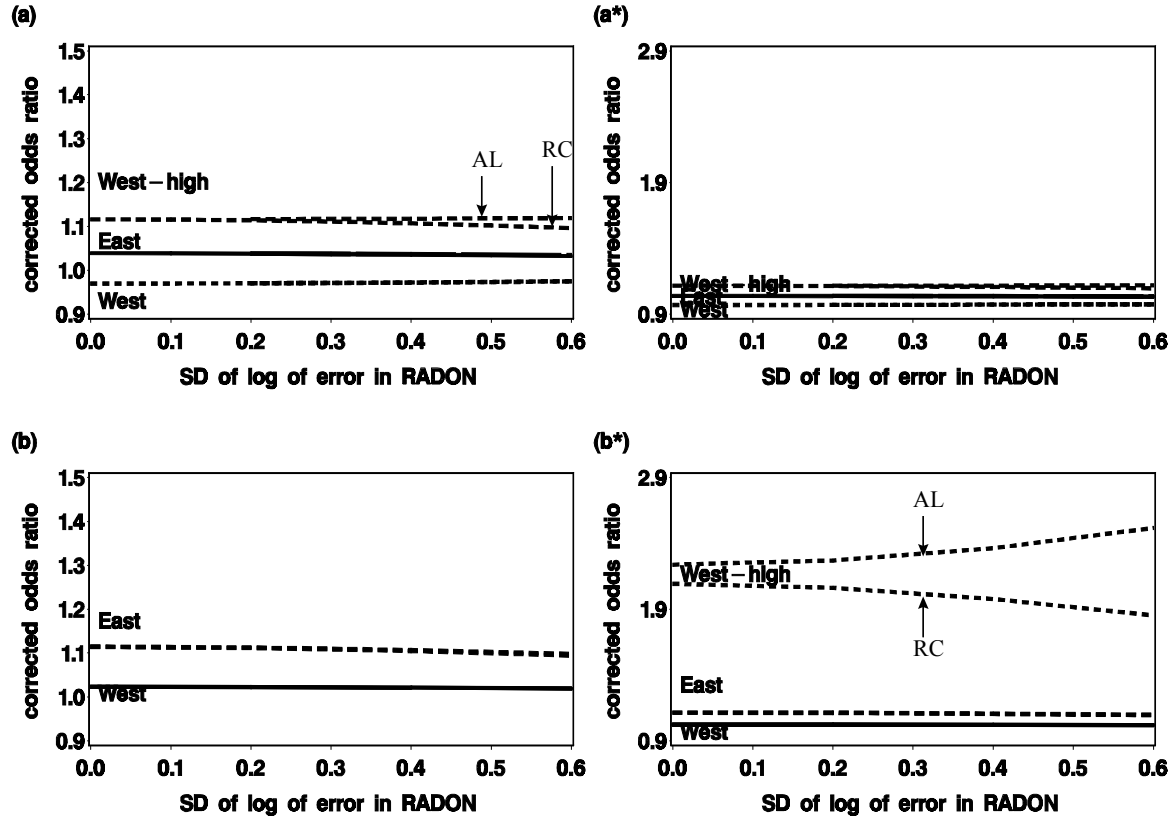

Figure 28: Analogous to Figure 25 (correcting for various sizes of multiplicative Berkson error in radon exposure assuming (a) no classical error, (b) a fixed classical error of size 0.4; (a) and (b*) show the same on a different scale of the vertical axis to fit West-High) comparing the approximate likelihood (AL) versus the regression calibration (RC) method of correction.*

4 DISCUSSION

It was our major objective to investigate the impact of uncertainties in the assessment of radon exposure and the most potent confounder, smoking, on the relative lung cancer risk estimates in the German case-control studies. These studies include a study from eastern parts of Germany (East study), a study from western parts of Germany (West study), and a subgroup of the West study from radon-prone areas (West-High) (Wichmann et al., 1998, 1999, Kreienbrock et al., 2001).

To address this objective, an appropriate method to correct the lung cancer risk estimates for errors in the predictors, the regression calibration method (Rosner et al., 1989), was analysed. Another correction method, an approximate likelihood method (Reeves et al., 1998), was investigated additionally. Since the error model is the most prominent model assumption for both of these correction methods, the components of the error in the two predictors, radon exposure and smoking, were identified and classified in order to establish a concise and feasible error model. A bivariate multiplicative error model combing classical and Berkson errors has been applied in order to cope with (1) the classical error in measuring radon exposure, (2) the classical error in assessing packyears, (3) the Berkson error by using radon exposure as a surrogate for the alpha dose, and (4) the Berkson error by using packyears as a surrogate for the lung dose from inhaled smoking carcinogens. An additive error model was also applied for comparison.

Overall, the estimates of relative lung cancer risk due to radon exposure, derived as odds ratio (OR) estimates, were found to be higher after accounting for the multiplicative classical error in radon exposure or/and packyears, when compared to the naive OR estimates (Naive OR estimates were 0.97, 1.12, 1.04 per 100 Bq/m³ increase in radon exposure for the West, the West-High, and the East study, see *Table 1*.). Hereby, the impact of correcting for error in radon exposure is remarkable for the three study populations, whereas the impact of correcting for errors in packyears is only relevant for the West study. Correcting for realistically sized errors by the regression calibration method, an error in radon exposure of 0.4 (explaining 50% of the radon exposure variance on the log-scale in the West study) and for an error in packyears of 0.5 (explaining 10% of the packyear variance on the log-scale) resulted in OR estimates of 1.12, 2.11, and 1.16 per 100 Bq/m³ in the West, the West-High, and the East

study, respectively. It was further found that one outlying observation (out of 3800) in the West study, which is also part of the West-High study, was very influential on the risk estimates (Naive OR without the outlier were 1.01 and 1.29 for the West and the West-High study, see *Table 12*). Without this outlier, correcting for the same errors as above yields OR estimates of 1.21 and 2.93 per 100 Bq/m³ (see *Table 12*) for the West and the West-High study, respectively. Applying the correction method by Reeves et al. (1998), the corrected OR estimate were similar (see *Figure 27*).

Berkson error was found to have less impact on the risk estimates than the classical error and slightly decreased the OR estimates, when applying regression calibration method (see *Figure 25* and *Table 14*). Correcting for multiplicative classical error in radon exposure of 0.4, a multiplicative classical error in packyears of 0.5, and a Berkson error from using radon exposure as a surrogate for the alpha dose of 0.6 (a 100% Berkson error, that is the error variance is as large as the observed radon exposure variance on the log-scale) by the regression calibration method, the OR estimates were 1.10, 1.87, 1.13 per 100 Bq/m³ for the West, the West-High, and the East study, respectively. Without the outlier in the West (and West-High study), the OR estimates were 1.17 and 2.45 per 100 Bq/m³ for the West and the West-High study. Applying the correction method by Reeves et al., the direction of the correction for Berkson error differed from the regression calibration method, resulting in slightly higher OR estimates (see *Figure 28*).

Modelling an additive classical error in radon exposure provided similar results for the West-High and the East study, but showed a qualitative difference for the full data analysis of the West study (see *Figure 23*), which, however, disappeared with exclusion of the outlier (see *Figure 24*). Neither additive Berkson error in radon exposure nor any Berkson error in the covariate had any impact on the OR estimates.

4.1 Error in assessing radon exposure (multiplicative classical error)

The error in assessing radon exposure was found to be well-modelled by a multiplicative classical error (see 3.1), and correcting for such error yielded higher OR estimates (see *Figure 20*a). For example, correcting for an error of size 0.4 yielded the OR estimate of 1.02, 2.09,

1.11 and for the West, the West-High, and the East study, respectively. For the East and the West-High study this is expected from theory. An increase in OR estimate after correction for multiplicative classical error was also found in the Swedish and the English study (Lagarde et al, 1997, Darby et al., 1998). However, for the West study, the increase in the OR and the reversal of the effect was a surprise (see below in 4.2), since the naive OR estimate in the West study was below unity.

The question of a statistically significant difference of the OR estimates from unity was not changed by the error correction in the West and the East study. However, the lower bound of the confidence intervals of the West-High study was larger than unity, correcting for errors of size 0.4. In 3.2.4, it was reasoned that the width of the confidence intervals widened with correction for classical error as the point estimates increased and that at the most borderline statistical [in-]significance, as in West-High, is changed by the error correction. However, the error size was assumed as known, and the confidence intervals would have further widened if the uncertainty in the error size had been additionally taken into account. For West-High, the statistical significance after correction should be regarded as it was before, i.e. borderline.

It can be concluded that accounting for the uncertainties in assessing radon exposure has remarkable impact on the quantification of the relative lung cancer risk due to increased radon exposure and that the risk is under-estimated if this uncertainty is ignored in the analysis.

4.2 Reversal of effect in the West study

The finding that the naive OR estimate for 100 Bq/m³ increase in radon exposure in the West study of 0.97 increased, instead of decreased, and even exceeded unity after accounting for a multiplicative classical error larger than 0.3 (SD on log-scale) was rather surprising. Generally, random non-differential error of the classical type is said to induce „bias towards the null hypothesis“: An OR estimate smaller than unity would be expected to decrease after correction; the „protective“ effect would be expected to intensify as described in 3.2.1. However, it was also mentioned in 3.2.1 that a mere tendency for the bias from multiplicative error towards the null could be stated, but that a dependence on data remained. In fact, correcting for an additive error resulted in decreasing OR estimate in West study.

This unforeseen finding was, however, plausible, considering the sensitivity of the OR estimate to one outlying observation. Excluding this observation from the West data set, the naive OR estimate was 1.01. Without this observation, the increase in the OR estimates when correcting for an increasing size of multiplicative classical error was now in line with the theory; the choice of error model, additive or multiplicative, did not change the results qualitatively (see *Figure 22* and *Figure 24*). *Figure 29* explains why correcting for multiplicative classical error in radon exposure is in fact dealing with this outlier problem.

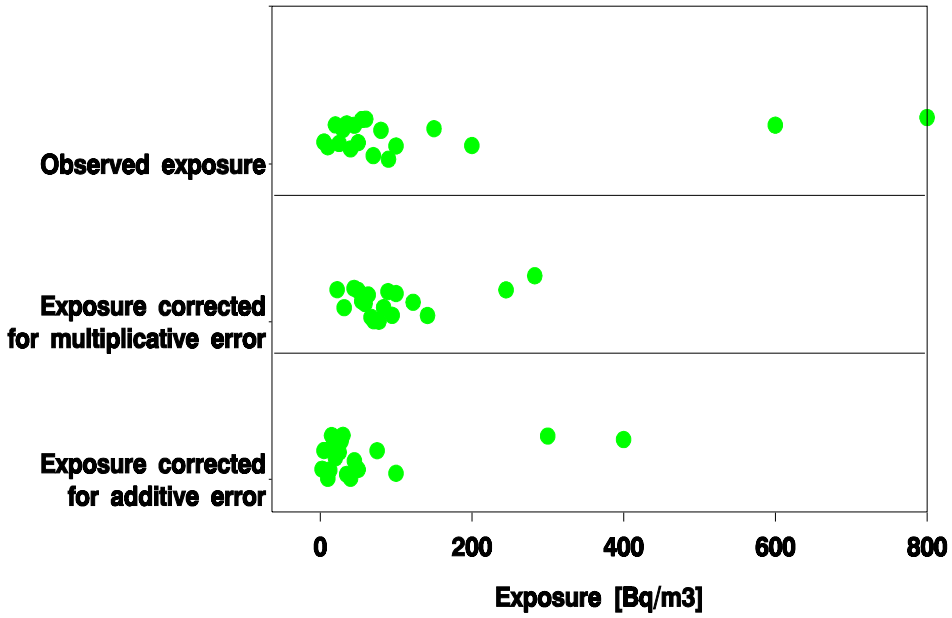

Figure 29: Fictitious data similar to radon exposure observed in the West study (although less observations are included for clarity): The upper frame depicts observed exposure. The largest and the second largest exposure is from a control, the third largest from a case. The middle and lower frame show these exposures corrected for multiplicative or additive error of the classical type, respectively.

It can be seen that correcting for multiplicative error not only draws extreme exposures closer to the middle of the data and thus reduces the absolute distance between observations more than correcting for additive error; it also diminishes their relative distance. Correcting for additive error simply involves the multiplication of the exposures by a constant smaller than unity, the reliability factor, and the relative distance between two exposures thus stays the same (see 3.2.1).

It is important to note that correcting for measurement error in exposure is not labelling extreme exposures as „false“ and replacing them with more „correct“ values; it merely makes the exposure values enter the model such as they can be expected to be true on average. The phenomenon that extreme observations tend to be less extreme when repeated has become quite famous as „regression towards the mean“ (Galton et al. 1886, Ibibarren et al. 1996).

4.3 Errors in assessing packyears (multiplicative classical error)

The error in assessing the smoking variable, packyears+1, was found to be mulitplicative classical (see 3.1). Correcting for such error in the confounder had practically no effect in the West-High study, increased the OR estimates in the East study, and most notably increased the estimates in the West study (see *Figure 20b*). The lack of effect in West-High is due to the lack of correlation between radon exposure and packyears in the West-High study area. In contrast, the increase in the OR estimates in the East and the West studies is due to the negative correlation between these risk factors. The observation that the impact of the error in packyears was larger on the West study estimate than on the East study estimate is due to the fact that correcting for errors in the covariate involves adjusting for the covariate differently, and that the adjustment for packyears already had greater impact 3.2.1 on the naive OR estimate in West when compared to East (see *Table 1*).

It was reasoned that the error in packyears were rather small, when compared to the error in radon exposure, and an error of size 0.5 seemed to be realistic (see 4.11 and Krewski, 2000). Correcting for such moderate error hardly increased the OR estimate in the West and the East study: corrected OR estimates were 0.99 and 1.05, respectively (instead of the naive estimates of 0.97 and 1.04). Correcting for large errors in packyears of size 1.0 increased the OR in the East, but more in the West study. In fact, it affected the West study to the extent that the OR coincided with the OR of the East study (corrected OR estimate of 1.09 for both) and approached the OR of West-High (The OR estimate of West-High, regardless of error in packyears, was 1.12). However, an error of such magnitude seemed to be exaggerated. Correcting for moderate error in the confounder had larger impact, if it was taken into account together with the error in radon exposure of size 0.4: The OR estimates were 1.12, 2.11, and 1.16 for the West, the West-High, and the East study (when compared to 1.02, 2.09, and 1.11

correcting only for the error in radon exposure of 0.4, see *Table 11*).

The OR estimate in the West study corrected for an error in packyears of 0.6 and larger was found to exceed unity. At first glance, this appeared to be as surprising as the fact, that the West OR estimate corrected for errors in radon exposure larger than 0.3 exceeded unity. However, a certain percentage of simulated case-control studies under the assumption of a true underlying radon risk of 1.12 per 100 Bq/m³ and additive or multiplicative classical error in the confounder showed a naive radon risk estimate below unity (data not shown). On the other hand, modelling classical error in the variable of primary interest, none of the simulated studies exhibited this effect. This indicates that the reversal of effect in the West study when correcting for error in radon exposure is specific for this study, whereas a reversal of effect when correcting for error in packyears is likely be observed also in other studies.

The width of the confidence intervals in the German studies was merely changed by errors in packyears, but the interval boundaries increased with the estimate. This resulted in a lower confidence bound larger than unity for the estimate in the East study after correction for an unrealistically large error in the smoking variable of 1.0 (see *Table 11*). However, taking this error into account together with an error in radon exposure of 0.4, this statistical significance disappears again. In 3.2.4, it was explained analytically that the correction of error in the confounder had usually no effect on the width of the confidence intervals.

It can be concluded that residual confounding due to adjusting for an imprecisely measured covariate, packyears, is neither a threat for the West-High study, nor does it greatly impact on the East study results; however, residual confounding has possibly deflated the estimate of relative lung cancer risk due to radon exposure in the West study. The usefulness of this sensitivity analysis, correcting for a range of error sizes, in understanding the magnitude of this problem is to be emphasised.

4.4 Impact of the correlation between radon exposure and smoking

In the West and the East study, the observed correlation between radon exposure and packyears (on the log-scale) is -0.06. In the presence of error in the predictor variables, the true correlation is larger. For an error of 0.4 in radon exposure and an error of 0.5 in

packyears, the corrected correlation was about -0.08 for both studies. The correlation is due to the fact that the study populations of these two studies included urban and rural areas, with the rural population being exposed to higher radon levels, but smoking less. In the West-High study area, the predictor variables were uncorrelated, which is due to the fact that the radon-prone areas are rather rural and do not exhibit the differences between the urban and the rural population as the full West or the East study.

Ignoring the correlation in the correction for multiplicative classical error in radon exposure (assuming no error in packyears) leaded to larger corrections and thus to higher OR estimates. This is plausible given the fact that taking the correlation into account is equivalent to making use of the information on radon exposure inherent in the correlated smoking variable; the effect of error in radon exposure is thus reduced, the corrections are smaller, the OR estimates less high. The correlation would have no effect if an additive error model was applied.

Error in the covariate only has impact in the presence of any correlation between the covariate and the predictor of primary interest and if such a correlation is accounted for. This is due to the fact that the only difference of the model correcting for errors in an uncorrelated confounder to the naive model is a slightly different adjustment for smoking, $Z_2^{\gamma_{22}}$ instead of Z_2 (see 3.2.2.2), which did not affect the OR estimates in the German radon studies, not even the West OR estimate. However, a confounder is per definition associated with the variable of primary interest, which is the primary reason for including the confounding variable. Thus, uncorrelated covariates in the exposure-disease model are rather the exemption than the rule. On the other hand, the difference between the model correcting for errors in a correlated confounder and the naive model is the adjusting for $Z_1^{\gamma_{21}} Z_2^{\gamma_{22}}$ instead of Z_2, which can have a clear impact, when adjusting for the confounder has already effected the raw OR estimates. It is interesting that even a small (true) correlation of -0.08 had such an impact on correcting both for classical error in radon exposure and for classical error in the confounder.

With regard to study design, a study region with the correlation between the variable of primary interest and the confounding variable being as small as possible appears to be preferable.

4.5 Residual confounding

Residual confounding refers to „the confounding bias, which remains after imperfect control for confounding“ (Brenner and Blettner, 1997). Although a consensus definition has not been reached (Kleinbaum et al., 1986, Last, 1995, Olsen, 1999, Breslow and Day, 1980), the underlying problem is of great relevance for epidemiological studies. Observational studies in general need to compensate for the lack of experimental control by statistical control of variables which might confound the results. Imperfect removal of confounding might be introduced by

(1) the error in assessing the confounding variable,
(2) the mis-specification of the functional relationship between confounder and disease, or
(3) the lack of other (latent, unmeasured) confounding variables.

The example below, to illustrate all three sources of residual confounding, is provided from the context of the radon studies.

The logistic model relating approximate relative lung cancer risk (RR) to radon exposure (RN) controlling for the confounder smoking by adjusting for the observed number of smoked packs per day (obsPACK) times the observed number of years of smoking (obsDUR), is

$$\log RR = \alpha + \beta_1 RN + \beta_2 (obsPACK * obsDUR) + \varepsilon \qquad (*)$$

(assuming that RN is observed without error). The perfect control of the confounder, smoking, may, however, be given by some other functional relationship of truePACK and trueDUR including other variables. Examples for such latent variables can be the age at start of smoking (AGE) or the time since exposure (TSE) to weight the smoked packs by the time elapsed since smoking them [Hauptmann et al. (2000) has shown that the amount of cigarettes smoked from two to 11 years before disease occurence is most predictive of lung cancer incidence.]

$$\log RR = \alpha + \beta_1 RN + \beta_2 f(truePACK , trueDUR, AGE, TSE...) + \varepsilon. \qquad (**)$$

Hence, using obsPACK and obsDUR instead of truePACK and trueDUR represents the first source of residual confounding, the error in assessing the confounding variables. Using the

multiplication of the two variables for the function f refers to the second source, the mis-specification of the model regarding the relationship of the confounder and the disease outcome. The lack of the variables AGE and TSE in the applied model (*) relates to the third source of residual confounding.

Residual confounding is often referred to in notes of caution when epidemiological studies are discussed. However, detailed methodological investigations are scarce and rather recent. Regarding source (1) of residual confounding, misclassification of discrete confounders were studied by Greenland (1980), Greenland and Robins (1985) and Savitz and Barón (1989). More recently, Marshall and Hastrup (1996) explored the impact of additive assessment error in continuous confounders on the effect of primary interest for rather large correlations (-0.3 to -0.7). Regarding source (2), Becher (1995) illustrated that adjusting for a continuous confounder was preferable to including categories of the confounder. Brenner and Blettner (1997) found that inclusion of a single linear term for the continuous confounder tended to be a reasonable choice for a broad variety of true confounder-risk associations. The only way in handling source (3) is thoughtful data collection and knowledge about the risk factors of the disease under study. In fact, the German radon studies involved extensive data collection on numerous aspects of smoking. The age at start of smoking showed no confounding effect on the risk from radon, and omitting it from the analysis induced thus no residual confounding. However, adjusting for packyears instead of for a more individual smoking history, taking the time since exposure (TSE) and the dependence of the effectiveness of cigarette smoking on TSE into consideration, has the potential of bearing residual confounding, since TSE has been shown to be of great importance (Hauptmann et al., 2000). Further, the existence of other latent smoking-related confounders cannot be ruled out completely.

With correcting for classical error in packyears and exploring the impact of Berkson error from using packyears as a surrogate for the lung dose, the direction and the dimension of residual confounding from source (1) was investigated in the German radon studies. Error in the packyears was found to be no problem for West-High, to have slight impact on the East study, and to mostly affect the West study. Berkson error from using packyears as a surrogate for lung dose had no direct impact on the radon risk (see 4.9). However, Berkson error also increases the imprecision of the smoking-effect estimate, the effect of which on the radon risk is not completely clear and should be subject to further research.

It can be concluded that identifying the sources of error in the assessment of the confounder and conducting a sensitivity analysis accounting for various sizes of errors in the assessment provides a valuable tool to explore the dimension of residual confounding from source (1) and is highly recommended to other investigators. However, further data analysis and methodological investigations regarding the sources of residual confounding, most importantly due to model mis-specification and other latent smoking-related variables, are necessary to understand the full dimension of the role of the potent confounder smoking for estimates of small risks.

4.6 Comparing West and West-High

As already stated, radon exposure and packyears are not correlated in the radon-prone areas of the West study, the West-High study, whereas there is a clear, but small negative correlation of these risk factors in the full West study population (see 4.4). Errors in packyears were found to attenuate OR estimates, if the packyears are negatively correlated with radon exposure. But such errors were found to have no impact in the case of no correlation (see 3.2.3.2). Further, the variance of radon exposure is smaller in the West study than in West-High (on the log-scale, but also on the original scale). And the impact of error in radon exposure is the larger, the smaller the radon exposure variance (see 3.2.2.2). It was thus an objective to clarify whether the bias from errors in the predictor variables differs between West and West-High to the extent that the different OR estimates, or a part of the difference, can be explained hereby (naive OR estimate of 0.97 versus 1.12 in West and West-High).

Table 16 contrasts the OR estimates corrected for various errors in the risk factors for the West study with the West-High study. It can be seen that classical error in radon exposure does not explain the difference between the naive OR estimates in the two study regions; on the opposite, such error even increases the difference. On the other hand, a moderate classical error in packyears explains some of the difference, a large error in packyears the majority of the difference. However, the error in radon exposure and the error in packyears need to be taken into account simultaneously. Then the difference, when compared to no correction, increased again. The difference of the OR estimates between West-High and West corrected for classical error in radon exposure was decreased by the additional correction for a moderate

error in smoking.

Correction for Berkson error from using radon exposure instead of alpha dose by the regression calibration method decreased the OR estimates particularly in West-High. A Berkson error of 0.6 would explain a part of the difference. However, the Berkson error needs to be taken into account together with the classical error in radon exposure and with a moderate classical error in packyears: Correction for the three errors simulaneously increased the difference between the OR estimate. Further, accounting for the Berkson error by the method by Reeves et al. (1998), the OR estimates increased in West-High and are basically untouched in West; thus the difference is not decreased, but increased, when the other correction method is used.

Table 16: *German radon studies: Comparison of odds ratio (OR) estimates in West and West-High after correction for various errors (applying the regression calibration method)*

	OR estimate West	*OR estimate West-High*
without error	0.97	1.12
multiplicative classical error:		
class error in radon exposure of 0.4	1.02	2.09
class error in smoking of 0.5	0.99	1.12
class error in smoking of 1.0	1.09	1.12
class error in smoking of 0.5	1.12	2.11
+ class error in radon exposure of 0.4		
multiplicative Berkson error:		
Berkson error of 0.6	0.97	1.10
Berkson error of 0.6	1.02	1.71
+ class error in radon exposure of 0.4		
Berkson error of 0.6	1.03	1.24
+ class error in smoking of 0.5		
Berkson error of 0.6	1.10	1.87
+ class error in radon exposure of 0.4		
+ class error in smoking of 0.5		

It is concluded that, assuming perfect assessment of radon exposure, moderate error in the packyears explains some, large error explains a majority of the difference of the naive OR estimates in the West-High and West study. However, acknowledging that no error in radon exposure is unrealistic, that an error of size 0.4 is very plausible, and that correcting for such error together with the error in packyears did not decrease the difference between West-High and West, it is further concluded that there need to be other mechanisms involved to satisfactorily explain the discrepancy between West-High and West.

4.7 Concordance between West and East

The extent of the bias from additive classical error depends on the exposure variance: The larger the variance, the smaller the bias. The same was shown for multiplicative error and the exposure variance on the log-scale by simulation studies (Heid et al., in press). There, it was concluded that the impact of classical error on the OR estimate in the West study would be larger than in the East study due to the smaller exposure variance. Note that these considerations did not include any covariates such as packyears.

Correcting for error in radon exposure in the German radon study data, larger corrections would have been found for the West study than for the East study to the extent that the two OR estimates would coincide, if there had been no correlation between radon exposure and packyears or if the correlation had been ignored (see *Figure 21*). However, accounting for the correlation, the OR estimates of the West and the East study showed about the same difference after the correction for errors in radon exposure as before (see *Figure 20a*).

On the other hand, the impact of correcting for error in packyears is larger in the West study than in the East study to the extent that the OR estimates coincide when correcting for a (unrealistically) large error of 1.0. This is plausible, since the exposure-disease model accounting for such error involves different adjustment for the confounder when compared to the naive model, and adjustment for smoking had already most strongly affected the naive estimates in West (see Table 1 and Section 3.2.3).

It is concluded that correcting for error in radon exposure or for a realistically sized error in

packyears (up to 0.5) does not reduce the difference between the OR estimates in the West and the East study. However, exclusion of one outlier (out of 3800) in the West study yielded almost identical OR estimates for the two studies and the similarity remained with correcting for error in the risk factors (see *Figure 22*).

4.8 The sensitivity on the choice of a multiplicative versus additive error

It was found that modelling an additive, instead of multiplicative, classical error in radon exposure changed the results in the West study qualitatively: The naive OR estimate of 0.97 further decreased (that is the „protective" effect increased) in the West study modelling an additive error, instead of exceeding unity with the multiplicative error (see *Figure 23*). This is in contrast to Lagarde et al. (1997), who suggested that the choice of additive versus multiplicative error had no major impact. However, with exclusion of the outlier in the West study, this major difference disappeared; correction for additive error then provided a similar picture as correction for multiplicative error (compare *Figure 24* with *Figure 22*).

For the East and the West-High study, the direction of the correction for additive classical error is the same as for multiplicative classical error, but the impact of the correction is smaller. Correcting for classical error in packyears, modelling an additive error makes no difference to modelling a multiplicative error.

4.9 Error from using external exposure as a surrogate for lung dose (multiplicative Berkson error) or „Does the lung dose really matter?"

The Berkson error was first described by J. Berkson (1950) and is the counterpart to the classical error. The classical error is the kind of error that first comes to mind when thinking about „measurement error". The Berkson error is less intuitive, but nonetheless frequently encountered in epidemiology. It arises particularly due to the use of a group average exposure in place of individual values. This is the case, when local area mean measurements are imputed for missing measurements (Reeves et al., 1998), when the entries of job-exposure matrices are used for exposure assessment instead of individual measurements, or when the

measurements of fixed monitors are used as predictor variable for a group of persons in the vicinity of these monitors. One example would be the use of the distance of one's residence to the nearest power station instead of measuring the electromagnetic field in the homes; other examples are using residential radon exposure as measurable surrogate of the truly experienced alpha dose, or using packyears as a surrogate for the lung dose of inhaled carcinogens released by smoking.

The exposure-disease models derived by regression calibration indicate that there is no bias on the radon risk from additive Berkson error in the predictor of primary interest and no bias from additive or multiplicative Berkson error in the covariate. It is thus concluded that there is no bias on the radon risk from controlling for packyears instead of the lung dose of inhaled smoking carcinogens. However, residual confounding by Berkson error in the confounder is possible due to the fact that the smoking effect is less precisely measured by using packyears instead of lung dose and thus, the estimate of this effect might, by chance, be not best suited to completely control for smoking (given the confounder-disease relation is modelled correctly and no other unmeasured confounders are missed!).

Correcting for multiplicative Berkson error from using radon exposure instead of alpha dose by the regression calibration method, the OR estimates were found to decrease, particularly in the West-High study: The bias from multiplicative Berkson error does not depend on exposure variance or on the correlation, but only on the effect size. Thus it is clear that the largest impact was found in the study with the largest effect. Note that the performance of the regression calibration method when correcting for large errors is not beyond any doubt. And it was found that the correction method by Reeves et al. (see *Figure 28*) leaded to different results when compared to the regression calibration method for large Berkson errors: The OR estimates slightly increased instead of decreased. For small Berkson errors, the difference between the two methods was negligible. The sensitivity analysis was performed for Berkson errors up to 100%, that is the proportion of the Berkson error variance is 100% of the observed predictor variance (on the log-scale); the true predictor variance is thus double the observed predictor variance. This is a fairly reasonable Berkson error size, but it might be larger than completely assure the well-performance of the correction methods, which should be subject to further research.

Generally, it can be stated that Berkson error hardly induces any bias on the risk estimates, but reduces the power of the study leading to wider confidence intervals. A real association might thus not be found. This is plausible, since the information on the predictor variable is cruder, when a group's value is used than when individual information is available. Correcting for Berkson error has thus only small effect on the risk estimates, if any. And there is neither an impact on the confidence intervals. It is, in fact, a drawback of the Berkson error that the loss of information cannot be corrected for, even if the error size was known. On the other hand, classical error can induce substantial bias and yield spuriously narrow confidence intervals. However, correction for the classical error not only results in approximately unbiased risk estimates, but also in corrected the confidence intervals. The erroneous information can be „subtracted". If the classical error size was precisely known, which is admittedly unrealistic, the power would be as good as in the case of no error in the predictor (see 3.2.4).

Every investigator has to face the question, whether to accept rather classical or Berkson error in the predictor, in the design phase of the study - a matter of defining the operationally defined predictor (see 3.1.1.4). In studies on the effect of electromagnetic radiation, for example, it is an on-going controversy, whether it is preferable to use the distance of one's residence to power stations as a surrogate for the individual exposure to electromagnetic fields or to directly measure electromagnetic radiation in each home: The first predictor being a crude, but fairly well measurable surrogate for the exposure of interest, thus a predictor with large Berkson, but small classical error; the latter being closer to the cause of the disease, thus a predictor with smaller Berkson error, but subject to substantial classical measurement error.

Due to the contrary effect of classical and Berkson error, the differentiation of each source of error in the predictor assessment process with regard to these two error types should be mandatory before applying the correction method (see 4.11).

4.10 Correction methods

The regression calibration method was applied to correct for errors in the risk factors, which is one of the most popular, most extensively studied, and most widely applied methods (for a review, see Thurigen et al. 2001). Furthermore, the method by Reeves was applied for

comparison, since this method was previously applied in the English radon study (Darby et al. 1998). Applying the method by Reeves yielded overall higher OR estimates than applying the regression calibration method: For classical error, the corrections were more extreme; for Berkson error, correction with the method by Reeves et al. resulted in larger OR estimates after correction, whereas correction by the regression calibration method yielded smaller OR estimates. Regarding the classical error correction, all conclusions made are valid no matter which of the two correction methods is applied. Regarding the Berkson error, the estimate-decreasing effect of the correction by regression calibration and the estimate-increasing effect of the correction by Reeves et al. should be interpreted with caution until more extensive performance evaluations are available (Küchenhoff et al., in preparation).

The interpretation of the models in a theoretical way in Section 3.2 was very valuable in many aspects: Being able to differentiate effects found in the German radon studies that were conclusive from the theory from effects that were inherent in the data was found to be important (see e.g. 4.2). Secondly, the theory helped with reducing the investigations on the data to the necessary: Correction for additive Berkson error in any predictor or for multiplicative Berkson error in the covariate were shown to have no impact and thus omitted from the data analysis plan. Thirdly, understanding the mechanism of the correction helped in understanding and trusting the reversal of the effect in the West study and the lack of impact from errors in packyears in the West-High study. Finally, it was attempted to shed some light into the „black box“ of measurement error correction to improve confidence into this method.

4.11 The error model

The sources of error in the predictor variables were identified meticulously and classified with care with regard to classical or Berkson type. All error components in assessing residential radon exposure (e.g. the error from the measuring process itself, the error from extrapolating a one-year measurement to concentrations of previous years) except the error from ignoring the occupancy time were found to be of the classical type. The error from ignoring the occupancy time and the error from using the residential radon exposure as a surrogate for the true alpha dose are of the Berkson type (see 3.1.3.1.1). The formulation of the error from using environmental exposure as a surrogate for the organ dose as Berkson type error was postulated

by many authors (Armstrong et al. 1989, Tosteson et al. 1985, Zeger et al. 2000). Further, the errors in assessing the packyears (e.g. false recall of number of cigarettes smoked per day or of the years of smoking) and the error from mis-specifying the relevant exposure time were found to be of the classical type, and the other errors from using the true packyears as a surrogate for the truly inhaled lung dose of smoking carcinogens (e.g. error from between-brand-variability or from between-person-differences of respiratory patterns) to be of the Berkson type (see 3.1.3.1.2). Thus a multiplicative error model for radon exposure and for smoking has been formulated allowing for classical and Berkson type errors. Such an error model for radon exposure was formulated by Reeves et al. 1998 and applied on the data from an English radon study (Darby et al. 1998), but with a different interpretation of the Berkson error, namely a Berkson error from imputing missing measurements by the local area mean. We modelled for a different Berkson error (using external exposure as a surrogate for lung dose) and extended the model to also allow for error in the confounder, packyears.

Formulating the bivariate error model is not revolutionary. However, applications of error correction in epidemiology so far are often restricted to univariate errors. On the other hand, theoretical investigations usually involve multi-dimensional error models using matrix notation, which are little transparent for the biomedical investigator. It was an objective to particularly illustrate the impact of the correlation between two predictors, without the full complexity of matrix notation (see 3.2.2.2 and 3.2.3.2).

Further, it is remarkable that the multiplicative error model does not appear more often in the literature, with many exposures in epidemiology being subject to multiplicative error due to the impact of proportions of concentrations in many ways. The circumstance that the multiplicative error model is additive on the log-scale helps with understanding the error. However, since the exposure enters the exposure-disease model on the original scale, differentiation of these models is absolutely necessary. The choice of multiplicative versus additive error was found to have great impact, particularly on the West study, in contrast to other authors (Lagarde et al. 1997). Thus, the assumption of the error model being multiplicative is crucial here. The evidence on the multiplicativity of the error in radon exposure as given by the three sets of replicate data (see 3.1.5) supported what was reported by others (Miles et al. 1990, Gunby et al. 1998, Wilcox et al., 1989, Darby et al., 1998, Lubin et al., 1995, Lagarde et al, 1998). A multiplicative error is also reasonable for the error in

packyears: The error is likely to be almost zero for non-smokers, small for moderate smokers and larger for heavy smokers. The multiplicative error structure implies such a constantly increasing deviation between true and assessed exposure with increasing exposure (Entering packyears+1 instead of packyears allows for a very small error among non-smokers.). In fact, OR estimates were adjusted for cigarette smoking by including packyears+1 on the original scale, instead of log(packyears+1) as in the primary analysis, in order to enable the modelling of a bivariate multiplicative error model rather than mixing multiplicative error in radon exposure with additive error in packyears.

The assumption of random, non-differential error in radon exposure is not challenged due to the objective nature of the measuring process. However, it cannot be completely ruled out that errors in smoking may be differential due to possible recall bias depending on case-control status, which, however, should be minimal in these studies due to in-person interviews by trained staff via standardised questionnaire. Also, systematic under-reporting or over-reporting of smoking is possible (Palmer et al. 1994). Only homoscedastic errors have been considered. Investigating the effect of heteroscedastic errors is beyond the scope of this work.

Regarding the size of the error in radon exposure, the replicate data from the intercomparison study showed an error from a between-measurement-variability (that is the deviation between measurements made under the same conditions) of about 0.1 for the German lab (see 3.1.5.3 and Kreienbrock et al. 1999). This is a minimal boundary for the size of classical error in the epidemiological setting: The error from between-measurement-variability in the German radon studies with radon concentrations measured in over 10000 homes by this laboratory is probably larger.

The year-by-year replicate data indicated that the error from the between-year-variability was much larger than the error from between-measurement-variability (estimate of 0.6, see 3.1.5.2). However, this result is to be regarded with caution: Firstly, the data was collected in houses not representative of the epidemiological situation with extremely high radon concentration (about 1000 times as high as in the German radon studies). Secondly, the estimate of both error components, the error from between-year-variability and the error from between-measurement-variability, was only 0.5, indicating that estimation of each error component's size and then drawing conclusions on the error from all sources together would

not give the right picture.

The estimate of the error from between-measurement-variability with the year-by-year replicate data was about 0.1 as with the intercomparison data. The estimate of the between-measurement-variability with the bedroom/living room data is 0.3 (see 3.1.5.1). This is quite high and not completely reliable due to the inherent differences between the rooms, which cannot be satisfactorily adjusted for, not only regarding floor size, but also for example different ventilation habits for bedrooms and living rooms. Attempts to adjust for the ventilation habit in the two rooms, a variable gathered in the German studies, resulted in indefinite likelihoods and uninterpretable (and thus not reported) results.

Table 17 provides an overview of the applied error models, the exposure distributions, the applied correction methods and corrected OR estimates of three articles on error correction in radon studies. In the English study (Darby et al. 1998), an error size of 0.48 was estimated by repeated measurements in homes over a series of years (including the error from between-measurement-variability and between-year-variability). An error of this size would explain 20% of the exposure variance (on the log-scale) observed in the English study, which is reasonable. However, such an error would explain 66% of the radon exposure variance (on the log scale) observed in the German West study, a rather large error. In the Swedish study, Monte Carlo simulations have imitated the error sources reflecting the Swedish radon concentrations and between-year-variations (Bäverstam et al. 1985) and have found an estimate for these errors of 0.32. Additionally including other sources of error, such as the error from between-owner-variability, between-subphase-variability and ignoring the occupancy time, resulted in an error size of 0.45. Lubin et al., 1995, simulated the radon pattern reflecting the U.S. situation and explored the effect of various error sizes of up to 1.09; such an error would explain over 300% of the German-West exposure variance (on the log-scale), which is impossible for a classical error as noted in 3.1.4. An error of up to 0.4 for radon exposure, which would explain 50% of the observed exposure variance on the log-scale in the West study (40% in the East study), seems to be realistic and was chosen for the German radon studies as upper boundary in order to stay conservative.

Table 17: Review of three articles on correction for errors in radon exposure and comparison with German studies: Applied error model, method of quantifying error size, method of error correction, exposure distribution, error size and corrected OR estimates (RC denotes regression calibration , AL denotes approximate likelihood by Reeves et al)

Author and study	*Error components accounted for*	*Error model*	*Method of quantifying error size*	*Method of error correction*
Darby et al., 1998, England (E)	(i) error from between-measurement-var., between-season-var., between-year-var.	mult. class.	Replicate data	AL
	(ii) imputing missings	mult. Berk		
Lagarde et al., 1998, Sweden (S)	(i) error from between-measurement-var., between-year-var.	mult. class.	Simulations	RC
	(ii) as (i) + error from between-subphase-var., between-owner-var., ignoring occupancy	mult.class	Simulations	RC
Lubin et al., 1995 , simulated U.S. situation (US)	error from between-measurement-var., between-room-var., between-year-var., between-subphase-var., between-owner-var., ignoring occupancy, imputing missings	mult.class	Sensitivity analysis	RC
Germany: West (G-W), East (G-E)	(i) as described in 3.1.2.1, (1.a) - (4.g)	mult. class	Sensitivity analysis	RC
	(ii) using surrogate for lung dose as in 3.1.2.1, (4.f) and (5.b)	mult. Berk.		

Table 17 (continued).

Author and study	*Distribution observed radon exposure: Mean and standard deviation (SD) on log-scale*	*Error size*			*naive OR estimates per 100 Bq/m³*	*corrected OR estimates per 100 Bq/m³*
		σ [(1)]	*COV* [(2)] ([(3)])	*%variance explained*		
Darby et al., 1998, England (E)	Cornwall: Mean of 3.24, SD of 0.91; Devon: Mean of 4.08, SD o f 1.05	0.48	51%	E: 20%, G-W: 66%	1.08 (0.97, 1.20)	1.12 (0.95, 1.33)
Lagarde et al., 1998, Sweden (S)		0.32	33% (35%)	G-W: 29% .	1.10 (1.01 - 0.22)	1.14 (0.02 - 0.31)
		0.45	47% (53%)	G-W: 58% .	as above	1.20 (0.04 - 0.44)
Lubin et al., 1995, simulated U.S. situation (US)	Mean of 3.21, SD of 1.13	0.40	42%	US: 12%, G-W: 50%	-	-
		0.69	78%	US: 37%, G-W: 140%		
		1.09	151%	US: 93%, G-W: 340%		
German-West (G-W),	G-W: Mean of 3.69, SD of 0.59	0.40	42%	G-W: 50% .	0.97 (0.78, 1.14)	1.02[(4)] (0.62, 1.68)
German-East (G-E)	G-E: Mean of 4.02, SD of 0.64	0.41	42%	G-E: 40% .	1.04 (0.96, 1.12)	1.13[(4)] (0.89, 1.43)

(1) Standard deviation of log of error.

(2) COV as standard deviation divided by mean on original scale, $\mathrm{COV} = \sqrt{\exp(\sigma) - 1}$

(3) COV as standard deviation divided by GM, $\mathrm{COV} = \exp(\sigma^2) \cdot \sqrt{\exp(\sigma^2) - 1}$

(4) accounting for the correlation between radon exposure and smoking

An error in packyears explaining 50% of the variance on the log-scale (SD on log-scale of 1.0), seemed to be exaggerated. Heuristically, an error of size 0.5 was motivated, that is an error explaining 12% of the observed packyear variance (on the log-scale) in the East study and 10% in West (see 3.1.4.3).

4.12 Other model assumptions

It is a matter of true concern that violations to model assumptions, as there are the model for the true exposure distribution, the exposure-disease-model and the error model, distort the results (Michels et al. 2001). The most prominent model assumption, the error model, is already discussed (see 4.12).

The true exposure distribution has been assumed to be reasonably lognormal. Radon exposure is generally described as such (Lubin et al. 1995, Lagarde et al. 1997, Reeves et al. 1998). The fact that the packyears are non-negative and a product (of the number of packs smoked per day and of the number of years of smoking) support the assumption of lognormally distributed packyears (The lognormal distribution is the limiting distribution for the product of distributions as the normal distribution is for the sum.). *Figure 1* supports the claim of rather lognormally distributed observed exposure. Although assumptions on the exposure distribution are required by most methods accounting for errors in exposure, among them the regression calibration method, these methods are relatively robust to violations of these assumptions (Caroll et al. 1995).

However, the distribution of packyears needs some special consideration due to zero exposures (Greenland et Poole, 1995): Non-smokers have zero packyears, which creates a problem in estimating the lognormal distribution parameters (The log of zero is undefined.). Imputing the value 0.05 for non-smokers is no option, because taking the logarithm leads to large negative numbers and unstable estimates of the lognormal distribution parameters. Since log(packyears+1) was entered into the model in the primary analysis, to which this analysis was aimed to be as close as possible without sacrificing the entry of the smoking variable on the original scale, packyears+1 was entered. It turned out that the lognormal distribution parameters were quite stable for packyears+1, and analogous computations only

including smokers, avoiding the problem of non-zero exposures, provided similar results (not reported).

The true exposure-disease-relationship was modelled by the logistic regression model. In occupational radon studies in uranium miners, a linear RR model has been applied (Samet al. 1989, Lubin et al. 1994), which is biologically plausible. The logistic regression model approximates the linear RR model for small exposures and was used to model the relationship between lung cancer and residential radon exposure by most investigators (e.g. Auvinen et al. 1996, Darby et al. 1998, Kreienbrock et al. 2001, Wichmann et al. 1999, Alavanja et al. 1999, Létourneau et al. 1994).

4.13 Notes of caution

There are a few comments to be made regarding potential limitations of our investigations. Many of these issues should be subject of future activities, but are beyond the scope of this work.

Firstly, the exposure-disease model accounting for errors in exposure would require further modification if the potential correlation of radon exposure with the binary variables in the model (categories of years since quitting smoking, indicator for smoking of other products or asbestos exposure) were accounted for or if even misclassification in these variables was considered. Secondly, detailed investigations regarding the performance of the regression calibration method and the method by Reeves are necessary. The performance of the regression calibration method particularly for large Berkson errors, also allowing for heteroscedasticity, are currently investigated by Küchenhoff et al. (in preparation)

Thirdly, the size of error in packyears and the non-differentiality of this error could be challenged. In fact, the sparcity of literature on error in assessing smoking variables via questionnaire is astonishing, but possibly due to the fact, that the risks from smoking are most prominent even without sophisticated methods. One recent work by Caraballo et al. (2001) comparing self-reported cigarette smoking with measured serum cotinine levels conclude that self-reports appear to be a very good indicator of actual smoking status. Additionally, the optimal treatment of zero exposure, particularly in lognormal distributions, not only with

regard to measurement error correction is subject of concern (Greenland and Poole, 1995). Modelling packyears with imputation of 0.05 packyears for non-smokers would result in wrong answers due to the inflated exposure variance. Modelling packyears+1 seemed to work well, but more detailed investigations would be helpful. Additionally, the size of error in radon exposure could not be estimated from the German radon study data and the supplied additional external data was not completely satisfactory. However, in an assessment process as complex as measuring radon exposure, it is extremely difficult to pinpoint an estimate, and a sensitivity analysis varying error size of a plausible range seemed to be most sensible.

Further, mixed error structures, that is multiplicative error in one predictor and additive error in another predictor can theoretically be accounted for by various correction methods. However, applications have so far not implemented such complex models. Applying the regression calibration method or the method by Reeves et al. would be problematic. Also, interaction between the predictors and the impact of errors in the predictors on the interaction need to be considered. Finally, only one of the sources of residual confounding was extensively explored, namely the impact of the error in assessing the confounding variable. However, further analysis regarding the German radon studies is necessary to grasp the full dimension of this issue.

5 SUMMARY

Case-control studies on lung cancer and residential radon exposure had been conducted in Germany. Relative risk estimates from primary analysis were now subject to accounting for uncertainties in radon exposure and in the most potent confounder smoking. The regression calibration method and an approximate maximum likelihood method were applied. The differentiation between classical error (from assessing radon exposure or packyears) or Berkson error (from using radon exposure instead of alpha dose, from using packyears instead of inhaled dose of smoking carcinogens) was of major importance in this analysis.

Estimates of relative lung cancer risk due to radon exposure were found to be higher after accounting for multiplicative classical error in radon exposure and packyears. Outliers in the data strongly influence risk estimates, but their impact is reduced, if classical error is accounted for. In one study, the influence of one outlier explained the particularly large risk-increasing impact of error correction. But also residual confounding due to adjusting for imprecisely measured packyears deflated the risk estimate in this study. It is interesting that the small correlation between radon exposure and packyears had this notable effect. On the other hand, classical errors in packyears had no large impact in the radon-prone study areas. Further, Berkson error did not induce substantial bias on the radon risk estimates, but possibly decreased the power to detect existing effects and inflated the confidence intervals.

It was concluded that such an analysis was extremely valuable to understand the impact of uncertainties in the risk factor of primary interest on the risk estimate under study and the potential for residual confounding by assessment errors in the smoking variable. Note that assuming some error in the risk factors is more realistic than assuming no error. With regard to study design, study regions with no correlation between the variable of primary interest and potential confounders are preferable.

However, the exact magnitude of the error could not be estimated based on the available data. Further investigations regarding residual confounding due to model mis-specification and latent smoking-related variables are necessary to grasp the full dimension of an important issue in epidemiology, i.e. the role of the outstanding confounder smoking for estimating small risks.

6 ZUSAMMENFASSUNG

In Deutschland waren Fall-Kontroll-Studien zu Lungenkrebs und Radon in Innenräumen durchgeführt worden. In der Schätzung des relativen Lungenkrebsrisikos wurden nun Unsicherheiten in der Radonexposition und im stärksten potentiellen „Confounder“ Rauchen berücksichtigt. Hierbei wurden die Methode der Regressionscalibrierung und eine approximative „Maximum Likelihood“ Methode angewandt. Die Unterscheidung zwischen klassischem Fehler (durch Erhebung der Radonexposition oder der Packungsjahre) und Berkson-Fehler (durch Verwendung von Radonexposition als Surrogat für Alpha-Dosis oder von Packungsjahren als Surrogat für Lungendosis durch inhalierte Karzinogene im Rauch) war von besonderer Bedeutung in dieser Analyse.

Die Risikoschätzer waren höher, wenn multiplikative klassische Messfehler in der Erhebung der Radonexposition und der Packungsjahre berücksichtigt wurden. Ausreißer in den Daten haben einen starken Einfluß auf Risikoschätzer, welcher jedoch durch Berücksichtigung klassischer Fehler reduziert wird. In einer Studie erklärte der Einfluß eines Ausreißers den besonders starken das Risiko erhöhenden Effekt der Fehlerkorrektur. Aber auch „Residual Confounding“ durch Adjustierung für ungenau erhobene Packungsjahre verringerte das beobachtete Risiko in dieser Studie. Es ist interessant, dass sich die kleine Korrelation zwischen Radonexposition und den Packungsjahren so stark auswirkte. In höher mit Radon belasteten Studiengebieten hatten klassische Fehler in den Packungsjahren keinen großen Effekt. Ferner bewirkte der Berkson-Fehler keine nennenswerte Verzerrung der Risikoschätzer, aber schmälerte die „Power“, um existierende Effekte zu erkennen.

Diese Analyse war extrem nützlich, um die Auswirkung von Messfehlern im primären Risikofaktor auf das zu untersuchende Risiko und das Potential von „Residual Confounding“ durch Fehler in der Rauchvariablen zu verstehen. Man halte sich vor Augen, dass die Annahme irgendeines Fehlers in den Risikofaktoren realistischer ist als die Annahme keines Fehlers. Bezüglich des Studiendesigns, ist eine Studienregion, wo die Primärexposition nicht mit dem potentiellen „Confounder“ korreliert ist, vorzuziehen.

Die genaue Fehlergröße konnte jedoch nicht aus den zur Verfügung stehenden Daten geschätzt werden. Außerdem sind weitere Untersuchungen hinsichtlich des „Residual Confounding“ durch Modell-Fehlspezifikation und latente Rauchvariable notwendig, um das volle Ausmaß eines wichtigen Punktes in der Epidemiologie zu verstehen, nämlich die Rolle des herausragenden „Confounder“ Rauchen für die Schätzung kleiner Risiken.

7 A GUIDE TO DEFINITIONS AND ABBREVIATIONS

7.1 List of definitions

Bequerel (Bq)	Number of radioactive decays per second
Bias	Difference between a true quantity and its estimate on average
Confounding bias	Confounding of the effect of primary interest by the effect of another risk factor
Information bias	Bias in the risk estimate due to error in the predictor of primary interest
Covariate	Predictor variable other than the predictor of primary interest (e.g. smoking in the German radon studies)
Case	Study participant having the disease under study at index date
Control	Study participant not having the disease under study at index date
Hospital control	Control recruited at the hospital
Population control	Control recruited via population/telephone registry
Confounder	Factor associated with the predictor of primary interest and the disease risk
Covariate	Risk factor included into the exposure-disease model additionally to the risk factor of primary interest
Equilibrium factor	Factor describing the equilibrium radon/radon daughters given the environmental conditions
Error	
Berkson type error	Error independent from measured predictor
Classical type error	Error independent from true predictor
Differential error	Error with different distributions for cases and controls
Heteroscedastic error	Error with varying structure or magnitude for the individuals under study
Homoscedastic error	Error with same structure and magnitude for each individual
Non-differential error	Error with the same distribution for cases and controls
Random error	Error with zero expectation
Systematic error	Error with non-zero expectation
Error model	Model linking true predictor and observed predictor
Exposure-disease model	Model linking disease outcome and risk factors
Instrumental data	Data on the measured predictor and on another measured predictor, which is associated with the true predictor (instrumental variable)
Matching	Method to cope with confounders: making the cases and controls comparable with respect to the distribution of the confounding variable by study design
Frequency matching	Choosing a group of controls with similar distribution to a group of cases with regard to the distribution of the confounder
Individual matching	Choosing one or more controls similar to one case with regard to the confounder

Multiple regression analysis	Analysis of study data based on a mathematical model linking disease outcome and several risk factors
Logistic regression	One type of multiple regression analysis linking the log of the odds ratio of disease linearly to the risk factors.
Packyears	Number of smoked cigarettes per day times number of years of smoking divided by 20 (assuming 20 cigarettes per pack)
Precision of an estimate	Variability of the estimate
Regression calibration	Method to account for errors in the predictors in risk estimation
Residual confounding	Confounding bias, which remains after imperfect control of confounding
Replicate data	Replicates of measured predictor
Stratification	Method to cope with confounders: analysing the data for groups with similar background risk
Surrogate for true predictor	A variable having no information about the disease status other than what is available in the true predictor of interest.
Validation data	Data on true and measured predictor
Validity of an estimate	Agreement of the estimate on average with the true underlying quantity

7.2 List of abbreviations

AL	Approximate Likelihood method (Reeves et al., 1998)
ANOVA	Analysis-of-variance
Bq/m³	Bequerel per cubic meter (Number of decays per second per cubic meter)
CI	Confidence interval
GM	Geometric mean
GSD	Geometric standard deviation
OR	Odds ratio
RC	Regression calibration method
RR	Relative risk
SD	Standard deviation

8 REFERENCES

Agricola, G. (ed.) 1928. German Translation by Schiffner, C De Re Metallica. VDI Verlag, Berlin.

Alavanja, M. C. R, Lubin, J. H., Mahaffey, J. A., Brownson, R. C. 1999. Residential Radon Exposure and Risk of Lung Cancer in Missouri. *Am J Public Health* 89:1042-1048.

Alanvanja, M.C.R., Brownson, R.C., Lubin, J.H., Berger, E., Chang, J., Boice, J.D. 1994. Residential radon exposure and lung cancer among non-smoking women. *J Natl Cancer Inst* 86:1829-1837.

Armstrong, B. G., Whittemore, A.S., Howe, G.R. 1989. Analysis of case-control data with covariate measurement error: Application to diet and colon cancer. *Stat Med* 8:1151-1163.

Armstrong, B. G. 1990. The effects of measurement errors on relative risk regression. *Am J Epidemiology* 132(6):1176-84

Auvinen, A., Mäkelinen, I., Hakama, M., Castrén, O., Pukkala, D., Reisbacka, H., Rytömaa, T. 1996. Indoor radon exposure and risk of lung cancer: a nested case-control study in Finland. *J Natl Cancer Inst* 88:966-972, Erratum 1998, *J Natl Cancer Inst* 90:401-402.

Axelson, O., AnderssonK., Desai, G., Fagerlung, I., Jansson, B., Karlsson, C., Wingren G. 1988. A case-referent study on lung cancer, indoor radon and active and passive smoking. *Scandinavian J Work Env Health* 14:286-292.

Axelson, O., Edling, C., Kling, H. 1979. Lung cancer and residency: A case-referent study on the possible impact of exposure to radon and its daughters in dwellings. *Scandinavian J Work Env Health* 5:10-15.

Bäverstam, U., Swedjemark, G.-A. 1991. Where are the errors when we estimate radon exposure in retrospect? *Rad Prot Dosimetry* 36(2/4):107-112.

Berkson, J., 1950. Are there two regressions? *J Am Stat Association* 45:164-180.

Blot, W.J., Xu, Z., Boice, J.D., Zhao, D., Stone, B.J., Sun, J., Jing, L., Faumeni, J.Fl. 1990. Indoor radon and lung cancer in China . *J Natl Cancer Inst* 82:1025-1030.

BMU, Bundesminister für Umwelt, Naturschutz und Reaktorsicherheit, ed. 1992. Die Exposition durch Radon und seine Zerfallsprodukte in Wohnungen in der Bundesrepublik Deutschland und deren Bewertung. *Veröffentlichungen der Strahlenschutzkommission, Band 19 (in German)*. Stuttgart, Germany: Gustav Fischer.

Boffetta, O., Kogevinas, M., Simonato, L., et al. 1995. Current perspectives on occupational cancer risks. *Int J Occup Environ Health* 1:315-325.

Breslow, N.E. and Cain, K.C. 1988. Logistic regression for two-stage case-control data. *Biometrika* 75:11-20.

Breslow, N. E., Day, N. E. 1980. *Statistical Methods in Cancer Research. Vol. I - The Analysis of Case-Control-Studies*, IARC.

Brüske-Hohlfeld, I., Möhner, M., Pohlablen, H., Ahrens, W., Bolm-Audorff, U., Kreienbrock, L., Kreuzer, M., Jahn, I.,k Wichmann, H.-E., Jöckel, K.-H. 2000. Occupational lung cancer risk for men in Germany: Results from a pooled case-control study. *Am J Epidemiol* 151:384-395.

Brüske-Hohlfeld, I., Wichmann, H.-E.. 1996. Lungenkrebs (in German). In: Wichmann, H.E., Schlipköter, H.W., Fülgraff, B. (ed.). *Handbuch der Umweltmedizin*, ecomed verlagsgesellschaft, Landsberg/Lech, V-1.3.4

Burkart, W., Weller-Mewe, E.M., Schneider, G. 1996. Ionisierende Strahlung (in German). In: Wichmann, H.E., Schlipköter, H.W., Fülgraff, B. (ed.). *Handbuch der Umweltmedizin*, ecomed verlagsgesellschaft, Landsberg/Lech, IV-2.3

Caroll, R. J. 1995. *Measurement Error in Nonlinear Models*. Chapman & Hall.

Caroll, R. J., Spiegelman C. H., Gordon Lan, K. K., Bailey K. T., Abbott, R. D. 1994. On errors-in-variables for binary regression models. *Biometrika* 71:19-25.

Carroll, R.J. and Wand, M.P. 1991. Semiparametric estimation on logistic measurement error models. *J Royal Stat Soc*, Series B 53:573-585.

Darby, S., Whitley, E., Silcocks, P., Thakrar, B., Green, M., Lomas, P., Miles, J., Reeves, G., Fearn, T., Doll, R. 1998. Risk of lung cancer associated with residential radon exposure in south-west England: a case-control study. *Brit J Cancer* 78(3):394-408.

Doll, R., Peto, R. 1981. *Causes of cancer*. Oxford University Press, Oxford.

Doll, R., Hill, A.B. 1952. A study of the aetiology of carcinoma of the lung. *Brit Med J* 2:1271-1286.

Environmental Protection Agency (EPA). 1992. A citizen's guide to radon (2nd ed.): The guide to protect your family from radon. EPA ANR. 464. Office of Public Awareness. Washington D.C.

Environmental Protection Agency (EPA). 1987. Radon reference manual. EPA Publ. Nr. 520/1-87-20. U.S. Govt. Print Off., Washington D.C.

Edling, C., Kling, H., Axelson, O. 1984. Radon in homes: A possible cause of lung cancer. *Scandinavian J Work Env Health* 10:25-34.

Field, R.W. 2001. A review of residential radon case-control epidemiologic studies performed in the United States. *Reviews Environ Health* 16(3):151-167.

Field, R.W., Steck, D.J., Smith, B.J., Brus, C.P., Neuberger, J.S., Fisher, E.L, et al. 2000. Residential radon gas exposure and lung cancer: The Iowa radon lung cancer study. *Am J Epidemiol* 151:1091-1102.

Fraser, G.E., Stram, D. 2001. Regression calibration in studies with correlated variables measured with error. *Am J Epidemiology* 154(9):836-844.

Fuller, W.A. 1988. *Measurement Error Models*. New York, Wiley.

Galton, F. 1886. Regression towards mediocrity in hereditary stature. *J Anthropol Inst* 15:246-63.

Gerken, M., Kreienbrock, L., Wellmann, J., Kreuzer, M., Wichmann, H. E. 2000. Models for retrospecitive quantification of indoor radon exposure in case-control studies. *Health Physics* 78(3):268-278.

Gerken, M., Kreienbrock, L. 1996. Radon (in German). In: Wichmann, H.E., Schlipköter, H.W., Fülgraff, B. (ed.). *Handbuch der Umweltmedizin*, ecomed verlagsgesellschaft, Landsberg/Lech, V-2.3.3

Gesell, T.F. 1983. Background atmospheric Rn-222 concentrations outdoors and indoors: A review. *Health Physics* 45:289-302.

Greenland, S. et Poole C. 1995. Interpretation and analysis of differential exposure variability and zero-exposure categories for continuous exposures. *Epidemiology* 6(3):326-328.

Gunby, J. A., Darby, S. C, Miles, J. C. H., Green, B. M. R, Cox, D. 1993. R. Factors affecting indoor radon concentration in the United Kingdom. *Health Physics* 64:2-12.

Hammond, E.C., Seidmann, H. 1980. Smoking and cancer in the United States. *Prev Med* 9:169-174.

Hammond, E.C., Horn, D. 1958. Smoking and death rates - report on forty-four months of follow-up of 187.783 mean. II Death rates by cause. *JAMA* 166:1294-1308.

Hardcastle, G.D. and Miles, J.C.H. 1996. Ageing and fading of alpha particle tracks in CR-39 exposed to air. *Radiation Protection Dosimetry* 67:295-298.

Heid, I.M., Küchenhoff, H., Wellmann, J., Gerken, M., Kreienbrock, L., Wichmann, H.E. In press. *Stat Med*.

International Agency for Research on Cancer (IARC). 1988. *Evaluation of the carcinogenic risk of chemicals to humans - Man-made mineral fibres and radon*, Vol.43. IARC Monograph, Lyon.

International Agency for Research on Cancer (IARC). 1986. *Evaluation of the carcinogenic risk of chemicals to humans - Tobacco*, Vol.38. IARC Monographs, Lyon.

International Commission on Radiological Protection (ICRP). 1994. Lung cancer risk from indoor exposures to radon daughters. ICRP Publ. Nr. 50. Pergamon Press, New York.

International Commission on Radiological Protection (ICRP). 1993. Protection against radon-222 at home and at work. ICRP Publ. Nr. 65. Pergamon Press, New York.

International Commission on Radiological Protection (ICRP). 1990. Recommendations of the International Commission on Radiological Protection. ICRP Publ. Nr. 60. Pergamon Press, New York.

International Commission on Radiological Protection (ICRP). 1987. Lung cancer risk from indoor exposures to radon daughters. ICRP Publ. Nr. 50. Pergamon Press, New York.

Ibibarren, C., Sharp, D., Burchfield, C.M., Ping, S., Dwyer, J. H. 1996. Association of Serum Total Cholesterol with Coronary Disease and All-Cause Mortality: Multivariate Correction for Bias Due to Measurement Error. *Am J Epidemiology* 143(5):463-471.

Jacobi, W. 1989. Dose to tissue and effective dose equivalent by inhalation of radon-222, radon-220 and their short-lived daughters. GSF-report S-626, Neuherberg.

Jacobi, W. 1964. The dose to the human respiratory tract by inhalation of short-lived 222Rn-and 220 Rn-decay products. *Health Physics* 10:1163-1174.

Jedrychowski, W., Becher, H., Wahrendorf, J., et al. 1990. A case-control study of lung cancer with special reference to the effect of air pollution in Poland. *J Epidemiol Community Health* 44:114-120.

Jöckel, K.H., Ahrens W., Wichmann, H.-E., et al. 1992. Occupational and environmental hazards associated with lung cancer. *Int J Epidemiol* 21:202-213.

Kemski, J., Siehl, A., Valdivia-Manchego, M. 1996. Kartierung des geogenen Radon-Potentials in der Bundesrepublik Deutschland. In: *Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit, ed. Forschung zum Problemkreis „Radon" - Vortragsmanuskripte des 9. Statusgespräches* (in German). Bonn, Germany.

Klotz, J.B., Petix, J.R., Zagraniski, R.T. 1989. Mortality of a residential cohort exposed to radon from industrially contaminated soil. *Am J Epidemiol* 129:1179-1186.

Kreienbrock, L., Poffijn, A., Tirmarche, M., Feider, M., Kies, A, Darby, S.C. 1999. Intercomparison of passive radon-detectors under field conditions in epidemiological studies. *Health Physics* 76(5):558-63.

Kreienbrock, L., Kreuzer, M., Gerken, M., Dingerkus, G., Wellmann, J., Keller, G., Wichmann, H. E. 2001. Case-control study on lung cancer and residential radon in West Germany. *Am J of Epidemiology* 153(1):42-52.

Kreuzer, M., Gerken, M., Kreienbrock, L., Welmann, J., Wichmann, H.-E. 2001. Lung cancer in lifetime non-smoking men - results of a case-control study in Germany. *Brit J Cancer* 84:134-140.

Kreuzer, M., Pohlabeln, H., Ahrens, W., et al. 1999. Occupational risk factors for lung cancer in young males. *Scand J Work Environ Health* 25:422-429.

Krewski. 2000. Personal Comment. Workshop on the European pooling exercise of studies on lung cancer and residential radon. Paris.

Küchenhoff, H., Thamerus, M., Heid, I.M., Kreienbrock, L. In preparation.

Kuha, J. 1994. Corrections for exposure measurement error in logistic regression models with an application to nutritional data. *Stat Med* 13:1135-1148.

Lagarde, F., Pershagen, G., Akerblom, G., Axelson, O., Bäverstam, U., Damber, L., Enflo, A., Svartengren, M., Swedjemark, G. A. 1997. Residential radon and lung cancer in Sweden: risk analysis accounting for random error in the exposure assessment. *Health Physics* 72:269-276.

Lees, R.E.M., Steele, R., Roberts, J.H. 1987. A case-control study of lung cancer relative to domestic radon exposure. *Int J Epidemiol* 16:7-12.

Létourneau, E. G., Krewski D., Choi N. W., Goddard, M. J., McGregor, R. G., Zielinski, J. M., Du, J. 1994. Case-control study of residential radon and lung cancer in Winnipeg, Manitoba, Canada. *Am J Epidemiology* 140:310-22.

Lubin J. H., Boice J.D. 1997. Lung cancer risk from residential radon: meta-analysis of eight epidemiologic studies. *J Natl Cancer Inst* 89:49-57.

Lubin, J. H., Boice, J. D. jr., Samet, J. M. 1995. Errors in exposure assessment, statistical power and the interpretation of residential radon studies. *Radiation Res* 144:329-341.

Lubin, J. H., Boice, J. D., Edling, C. H., Hornung, R., Howe, G., Kunz, E., Kusiak, A., Morrison, H. I., Radford, E. P., Samet, J. M., Tirmarche, M., Woodward, A., Xiang, Y. S., Pierce, D. A. 1994. Radon and lung cancer risk: a joint analysis of 11 underground miner studies. Rockville, MD: US National Institutes of Health, *NIH publication* no. 94-3644.

Lucas and Woodward. 1964. Effect of long decay chains on the counting statistics of radium and radon, *J Applied Physics* 452-456.

Martignoni, K., Burkart, W. 1996. Ableitung von Grenzwerten (Umweltstandards) - Strahlung (in German). In: Wichmann, H.E., Schlipköter, H.W., Fülgraff, B. (ed.) *Handbuch der Umweltmedizin*, ecomed verlagsgesellschaft, Landsberg/Lech, III-1.2.

Michel, J. 1987. Sources. In:Cothern, C.R., Smith, J.E. (ed.) *Environmental radon*. Plenum Press, New York.

Michels, K. B. 2001. A renaissance for measurement error. *Int J Epidemiology* 30:421-422.

Müller, P., Roeder, K. 1997. A Baysian semiparametric model for case-control studies with errors in variables. *Biometrika* 84:523-537.

Mumford, J.l., He., X.Z., Chapman, R.S., et al. 1987. Lung cancer and indoor air pollution in Xuan Wie China. *Science* 253:217-235.

National Academy of Sciences (NAS), National Research Council. 1994. *Health effects of exposure to radon: Time for reassessment?. BEIR VI Report of the Committee on the Biological Effects of Ionizing Radiation* National Academy Press. Washington, D.C.

National Academy of Sciences (NAS), National Research Council. 1991. *Comparative dosimetry of radon in mines and homes.* National Academy Press. Washington, D.C.

National Academy of Sciences (NAS), National Research Council. 1988. *Health risks of radon and other internally deposited alpha-emitters. BEIR IV Report of the Committee on the Biological Effects of Ionizing Radiation* National Academy Press. Washington, D.C.

National Academy of Sciences (NAS), National Research Council. 1980. *The effects on populations of exposure to low levels of ionizing radiation. BEIR III Report of the Committee on the Biological Effects of Ionizing Radiation* National Academy Press. Washington, D.C.

Ooi, W.L., Elston R.C., Chen, V.W., et al. 1986. Increased familial risk for lung cancer. *J Natl Cancer Inst* 76:217-222.

Palmer, R. F., Dwyer, J. H. 1994. A measurement error model of adolescent smoking. *Addictive Behaviors* 19(5):447-489.

Pinel, J., Feran, T., Darby, S.C., and Miles, J.C.H. 1995. Seasonal correction factors for indoor radon measurements in the United Kingdom. *Rad Prot Dosimetry* 58:127-132.

Pershagen, G. 1993. Personal comment.

Pershagen, G., Axelson, O., Clavensjö, B., Damber, L., Desai, G., Enflo, A., Lagarde, F., Mellander, H., Svartengren, M., Swedjemark, G.A., Akerblom, G. 1994. Residential radon exposure and lung cancer in Sweden. *New England J Med* 330:159-164.

Pershagen, G., Liang, L., Hrubec, Z., Svensson, C., Boice, Jr., J. D. 1992. Residential radon exposure and lung cancer in swedish women. *Health Physics* 63(2):179-186.

Pierce, D.A., Stram, D.O., Vaeth, M. Allowing for random errors in radiation dose estimates for the atomic bomb survivor data. 1990. *Radiat. Res.* 123: 275-284.

Poffign, A., Tirmarche, M., Kreienbrock, L., Kayser, B., Darby, S.C. 1992. Radon and lung cancer: Protocol and procedures of the multi-centre studies in the Ardennes-Eifel region, Brittany, and the Massiv Central. Rad Prot Dos 45, supplement 1 / 4, 651-56.

Reeves, G. K., Cox D. R., Darby S. C., Whitley E. 1998. Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Stat Med* 17:2157-77.

Richardson, S. and Gilks, W.R. 1993a. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *Am J Epidemiology* 138:430-432.

Richardson, S. and Gilks, W.R. 1993b. Conditional independence models for epidemiological studies with covariate measurement error. *Stat Med* 12:1703-1722.

Rosner, B. et Gore, R. 2001. Measurement error correction in nutritional epidemiology based on individual foods, with application to the relation of diet to breast cancer. *Am J Epidemiol* 154(9):827-835.

Rosner, B.A. 1996. Measurement error models for ordinary exposure variables measured with error. *Stat Med* 15:293-303.

Rosner, B., Willett, W. C. and Spiegelman, D. 1989. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med* 8:1051-1069.

Ruosteenoja, E.., Mäkeläinen, I., Rytömaa, T., Hakulinen, T., Hakama, M. 1996. Radon and lung cancer in Finland. *Health Physics* 71(2):185-189.

Samet, J. M. 1989. Radon and lung cancer. *J Natl Cancer Inst* 81:745-757.

Sandler D.P., Weinberg, C.R., Archer, V.E., Rothney-Kozlak, L., Bishop, M., Lyon J.E., Stolwijk, J. 1999. A case-control study in Connecticut and Utah. *Proceedings of the American Statistical Association*. Conferecnce on Radiation and Health: Indoor radon and lung cancer risk. *Radiat Res* 151:103-105.

Schoenberg, J. B., Klotz, J.B., Wilcox H.B., Nicholls, G. P., Gil-del-Real, M. T., Stemhagen, A., Mason, T. J. 1990. Case-control study of residential radon and lung cancer among New Jersey women. *Cancer Res* 50:6250-6254.

Schoenberg, J.B., Klotz, J.B., Wilcox, H.B., Szmaciaz, S.F. 1992. A case-control study of radon and lung cancer among New Jersey women. *Twenty-Ninth Hanford Symposium on health and the Environment*. Indoor radon and lung cancer: Reality or Myth? United States Department of Energy and Battelle, Pacific Northwest Laboratories (sponsors). Columbus, Richland, Washington, USA: Battelle Press, 905-918.

Simonato, L., Boffetta, P., Kogevinas, M. 1996. Epidemiological aspects of cancer risk associated with exposure in the occupational environment. *Med Lav* 87:5-15.

Simpson, S.G., Comstock, G.W. 1983. Lung cancer and housing characteristics. *Arch Env Health* 38:248-251.

Stefanski, L.A., and Carroll, R.J. 1987. Conditional scores and optimal scores in generalized linear models. *Annals of Statistics* 13:1335-1351.

Steindorf, K., Lubin J.H., Wichmann, H.E., Becher, H. 1995. Lung cancer deaths attributable to indoor radon exposure in West Germany. *Int J Epidemiology* 24:485-92.

Strahlenschutzkommission (SSK). 1992. Die Eposition durch radon und seine Zerfallsprodukte in Wohnungen in der Bundesrepublik Deutschland und deren Bewertung (in German). Veröffentlichung der Strahlenschutzkommission Bd. 19, G. Fischer Verlag, Stuttgart.

Strahlenschutzkommission (SSK). 1988. Strahlenschutzgrundsätze zur Begrenzung der Strahlenexposition der Bevölkerung durch Radon und seine Zerfallsproduckte (in German). BMU, Bundesanzeiger Nr. 208, Bonn.

Strahlenschutzkommission (SSK). 1985. Strahlenexposition und mögliches Lungenkrebsrisiko durch Inhalation von Radon-Zerfallsprodukte in Häusern (in German). In: Empfehlung der Strahlenschutzkommission 1985/1986 Bd. 6, G. Fischer Verlag, Stuttgart.

Svensson, C., Eklund, G., Pershagen, G. 1987. Indoor exposure to radon from the ground and bronchial cancer in women. *Int Arch Occupat Env Health* 59:123-131.

Tosteson, T.D., Stefanski, L.A., Schafer, D.W. 1989. A measurement-error model for binary and ordinal regression. *Stat Med* 8:1139-1147.

Thürigen, D., Spiegelmann, D., Blettner, M., Heuer, C., Brenner, H. 2000. Measurement error correction using validation data: a review of methods and their applicability in case-control studies. 2000. *Stat Methods Med Res* 9:447-474.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). 1986. Genetic and somatic effects of ionizing radiation. Report to the National Assembly, with annexes. United Nations; New York.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). 1982. Ionizing radiation: sources and biological effects. Report to the General Assembly, with annexes. United Nations; New York.

United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). 1977. Sources and effects of ionizing radiation. United Nations; New York.

Vineis, P., Simonato, L. 1991. Proportion of lung and bladder cancers in males resulting from occupation: a systematic approach. *Arch Environ Health* 45:6-15.

Wichmann, H.E., Gerken, M., Wellmann, J., Kreuzer, M., Kreienbrock, L., Keller, G., Wölke, G., Heinrich, J. 1999. Lungenkrebsrisiko durch Radon in der Bundesrepublik Deutschland (Ost) - Thüringen und Sachsen (in German). *Fortschritte in der Umweltmedizin*. ecomed verlagsgesellschaft.

Wichmann, H.E., Kreienbrock, L., Kreuzer, M., Gerken, M., Dingerkus, G., Wellmann, J., Keller, G. 1998. Lungenkrebsrisiko durch Radon in der Bundesrepublik Deutschland (West) (in German). *Fortschritte in der Umweltmedizin*. ecomed verlagsgesellschaft.

Wichmann, H.E, Kreienbrock, L. 1996. Umweltepidemiologie (in German). In: Wichmann, H.E., Schlipköter, H.W., Fülgraff, B. (ed.) *Handbuch der Umweltmedizin*, ecomed verlagsgesellschaft, Landsberg/Lech, III-1.2.

Wrixon, A.D., Green, B.M.R., Lomas, P.R.M, Miles, J.C.H., Cliff, K.D., Francis, E.A. Driscoll, C.M.H., James, A.C., and O'Riordan, M.X. 1988. Natural radiation exposure in UK dwellings. *NRPB R*-190.

Zeger, S.L., Thomas, D., Dominici, F., Samet, J.M., Schwartz, J., Dockery, D., Cohen, A. 2000. Exposure measurement error in time-series studies of air pollution: Concepts and consequences. *Environ Health Perspectives* 108(5):419-426.

9 ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Dr. Dr. H.-E. Wichmann, Chair of Epidemiology, Institute of Medical Information Processing, Biometry and Epidemiology of the LMU, Munich, and Director of the Institute of Epidemiology of the GSF-National Research Centre for Environment and Health, Neuherberg, for motivating this work and his competent and benevolent supervision. I am grateful for the opportunity to work with the data of the German radon studies and for the possibility to participate at national and international conferences. Further, I would like to thank Prof. Dr. L. Kreienbrock, Director of the Institute of Biometry, Epidemiology and Information Processing of the School of Veterinary Medicine, Hanover, who has been very helpful with advice in various aspects of this work and who has made the data set on the laboratory intercomparison available. I am also grateful to R. Lehmann from the Federal Office of Radiation Protection; Berlin, who has kindly provided the data set of the year-by-year replicate radon measurements. I want to further thank Dr. J. Miles from the National Radiation and Protection Board, U.K., who has helped me a great deal with literature on errors in the radon concentration measuring process. I also want to state my gratitude to Prof. S. Darby from the Clinical Trial Service Unit of the Radcliffe Infirmary, Oxford, U.K., whose advice has been very valuable. I have also profited a lot from Prof. R. Carroll's comments at a conference in Munich, and from Prof. D. Krewski's feedback at a workshop in Paris. Special thanks to PD Dr. H. Küchenhoff from the Statistical Faculty of the LMU, Munich, for providing statistical supervision and often personal encouragement. Most of what I know about measurement error correction, I have learned from him.

Many thanks to all collaborators, who have contributed during seven years of data collection and preparation to the success of these large epidemiological studies. I acknowledge the huge amount of time and thoughts by each of them. I am very grateful to my colleagues from the Radon team for numerous discussions regarding radon, lung cancer and uncertainties in the exposure assessment, particularly the Drs. I. Brüske-Hohlfeld, M. Gerken, J. Wellmann, and A. Schaffrath Rosario. Many thanks to Tania Schwiewagner for searching for and copying a whole bunch of articles. For personal feedback and more than a few hot chocolates, I want to thank Dr. I. Meyer, who has been a precious friend and colleague. For review regarding the correct use of the English language, I thank Dr. M. Coles. Further, I want to acknowledge my gratitude to Prof. J. R. O'Fallon, former Director of the Cancer Centre Statistics of the Mayo Clinic in Rochester, Minnesota. She is living proof that successful research and having a family is not a paradox and, in fact, started my inclination towards biostatistics.

I dedicate this thesis to my daughters, Marina and Julia, who are the delight of my life, to my parents, Erika and Remigius Heid, and to my dearest husband, PD Dr. med. A. Luchner, who have never failed with encouragement, caring and advice.

CURRICULUM VITAE

Name:	Iris Heid
Date and Place of Birth:	January 16th, 1968, Munich, Germany

Academic Education:

1987-1994	Major Mathematics, Minor Physics at Ludwig Maximilian University (LMU), Munich, Germany
1994	Dipl. Math. Univ. (M.Sc.)

Chronology of Employment:

intermittendly during 1987-1994	Research Assistant (as student) at Institute of Mathematics, LMU Munich, Germany at Max-Planck-Institute, Garching, Germany and at GSF-Institute, Neuherberg, Germany
March 1994 - April 1994	Field work for epidemiological study MONICA, University of Regensburg, Germany
July 1994	Travel to U.S.A., application for H-I visum
April 1995 - Aug. 1996	Data Analyst and Junior Statistician (evaluation of clinical phase II and III trials) Cancer Centre Statistics, Mayo Clinic, Rochester, MN, USA
Oct. 1996 - March 1998	Systems Analyst (application software for paper industry, consulting and realisation) Mathematische Analysen und Systeme GmbH, Regensburg, Germany
since April 1998	Scientific Staff Member Chair of Epidemiologie (Prof. Dr. Dr. H.-E. Wichmann), Institute of Medical Information Processing, Biometry and Epidemiology, LMU Munich, Germany
April 1999- Sept. 1999	Maternity leave
since October 1999	PhD-Student, Faculty of Medicine, LMU at Munich, Germany
Mai 2002 - Sept. 2002	Maternity leave
October 9th, 2002	Date of oral examination