UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
Corso di Dottorato di Ricerca in Scienze Statistiche
Ciclo XXXVI

Advances in Control Risk Regression

Coordinatore del Corso: Prof. Nicola Sartori

Supervisore: Prof. Annamaria Guolo

Co-supervisore: Prof. ...

Dottorando/a: Thien Phuc Tran

January 10, 2024

Abstract

Meta-analysis is a long-established and widespread tool to summarize, aggregate and combine results from independent studies about the same issue of interest. As independent studies included in a meta-analysis differ in many ways, properly accounting for between-study heterogeneity is a relevant goal. The traditional meta-analysis model has a random-effect formulation, where between-study heterogeneity is accounted for through a variance component. When available, study-specific covariates can be inserted in the model, in order to better explain heterogeneity due to, for example, differences in studies' design and characteristics of participants. The resulting model is a meta-regression model, where additional covariates represent information summarized at the study level, and thus can be affected by aggregation error.

This thesis focuses on control risk regression, which is an example of meta-regression used in medical investigations to evaluate the effectiveness of a treatment in clinical trials comparing a treatment group and a control group. Control risk regression is characterized by the inclusion of a summarized measure of risk for the subjects in the control condition (control rate) as a covariate in the meta-regression model. Such a covariate represents a proxy for the underlying risk, that is, the measure of risk at the population level useful to describe unmeasurable sources of heterogeneity associated to a disease, as, for example, the severity of illness. Control rate is thus affected by measurement error. An appropriate analysis should correct for the presence of errors in order to provide reliable inference.

The thesis focuses on two extensions of the classical control risk regression model.

First, the model is extended to include additional study-specific covariates other than the control rate, as a way to provide a more accurate explanation of the heterogeneity. Likelihood-based inference is carried out by including measurement error corrections to prevent biases due to error in the control rate and errors in the additional covariates.

Attention is paid to an approximate normal specification of the measurement error structure as well as to an exact, and more computationally involved, specification. The lack of information about within-study covariances between risk measures and the covariate components is overcome by deriving explicit expressions using Taylor expansion based on study-level covariate subgroup summary information. As an alternative, a more efficient solution based on a pseudo-likelihood solution is developed, under a working independence assumption between the observed error-prone measures. The methods are evaluated in a series of simulation studies under different specification for the sample size, the between-study heterogeneity, as well as the underlying risk distribution. The methods are applied to real meta-analyses about the association between COVID-19 and schizophrenia, and the association between COVID-19 and myocardial injury.

A second extension of the classical control risk regression model intends to modify the linear relationship between the true treatment risk and the true control risk, which is motivated by convenience, although it is not always reasonable. The proposal is a U-shaped relationship between the risk measures, in this way allowing to describe treatments which have a positive effect and a negative effect. The price to pay is in terms of computational issues, since the likelihood function loses a closed-form expression, even under an approximate normal measurement error specification. The method is evaluated in a series of simulation studies, involving scenarios of different sample sizes and between-study heterogeneity, absence or presence of linear/quadratic relationships between the risk measures. The approach is applied to a meta-analysis about the association between diabetes and Parkinson's disease and to re-analyze the data about the association between COVID-19 and myocardial injury used in the first part of the thesis.

Sommario

La meta-analisi è uno strumento consolidato e diffuso da tempo per riassumere, aggregare e combinare i risultati di studi indipendenti riferiti allo stesso oggetto di interesse. Poiché gli studi indipendenti inclusi in una meta-analisi differiscono in molti modi, un obiettivo rilevante è tenere adeguatamente conto dell'eterogeneità tra gli studi. Il modello tradizionale di meta-analisi è un modello ad effetti casuali in cui l'eterogeneità tra studi viene tenuta in considerazione attraverso una componente di varianza. E' possibile inserire nel modello covariate studio-specifiche al fine di fornire una migliore spiegazione della eterogeneità, la quale può essere dovuta, ad esempio, alle differenze nel disegno degli studi e nelle caratteristiche dei partecipanti. Il modello risultante è un modello di meta-regressione, in cui le covariate sono informazioni riassuntive a livello di studio e quindi possono essere soggette da errori di aggregazione.

Questa tesi si concentra sulla regressione con rischio per il gruppo di controllo, la quale è un esempio di meta-regressione utilizzata prevalentemente in Medicina per valutare l'efficacia di un trattamento negli studi clinici che confrontano un gruppo di trattati e un gruppo di controllo. La regressione con rischio per il gruppo di controllo è caratterizzata dall'inclusione di una misura aggregata di rischio per i soggetti nella condizione di controllo (tasso di controllo) che figura come covariata nel modello di meta-regressione. Tale covariata è una rappresentazione del rischio di base, cioè una misura del rischio a livello di popolazione utile per descrivere fonti non misurabili di eterogeneità associate ad una malattia, come, ad esempio, la gravità della malattia. Per tale ragione il tasso di controllo è una quantità affetta da errore di misura. Un'analisi appropriata deve correggere per la presenza di errori al fine di fornire risultati inferenziali affidabili.

La tesi si concentra su due estensioni del classico modello di regressione con rischio per il gruppo di controllo.

Innanzitutto, il modello viene esteso per includere ulteriori covariate studio-specifiche diverse dal tasso di controllo, in modo da fornire una spiegazione più accurata dell'eterogeneità. L'inferenza basata sulla verosimiglianza viene effettuata includendo correzioni degli errori di misurazione per prevenire distorsioni dovute a errori nel tasso di controllo ed errori nelle covariate aggiuntive. Si considera sia un modello approssimato normale per l'errore di misura, sia un modello esatto, sebbene più complesso dal punto di vista computazionale. Le covarianze interne agli studi tra le misure di rischio e le covariate, solitamente non disponibili, vengono esplicitamente derivate tramite un'espansione di Taylor che sfrutta le informazioni sommarie relative a sottogruppi delle covariate a livello di studio. In alternativa, viene sviluppata una soluzione più efficiente basata sulle pseudo-verosimiglianze, sotto l'assunzione di indipendenza tra le quantità soggette a errore di misurazione. I metodi proposti vengono valutati in una serie di studi di simulazione al variare della dimensione del campione, dell'eterogeneità tra gli studi e della distribuzione del rischio sottostante. I metodi vengono applicati a meta-analisi reali sull'associazione tra COVID-19 e schizofrenia e sull'associazione tra COVID-19 e danno miocardico.

Una seconda estensione del classico modello di regressione con rischio per il gruppo di controllo prevede la modifica della relazione lineare tra il vero rischio nella condizione di trattamento e il vero rischio nella condizione di controllo, che è solitamente adottata per motivi di convenienza, sebbene non sia sempre una scelta ragionevole. La proposta è una relazione a forma di U tra le misure di rischio, che permette di descrivere trattamenti che hanno un effetto positivo e un effetto negativo. La scelta ha un costo in termini computazionali, poiché la funzione di verosimiglianza associata non è più in forma chiusa, anche nel caso di errori di misura assunti approssimativamente normalie. Il metodo viene valutato in una serie di studi di simulazione, che coinvolgono scenari con diverse dimensioni del campione ed eterogeneità tra studi, sia in assenza sia in presenza di relazioni lineari/quadratiche tra le misure di rischio. L'approccio viene applicato ad una meta-analisi sull'associazione tra diabete e malattia di Parkinson e per analizzare nuovamente i dati sull'associazione tra COVID-19 e danno miocardico utilizzati nella prima parte della tesi.

To my mom, the strongest woman I know.

Acknowledgements

I wish to show my appreciation to my advisor, Prof. Annamaria Guolo. She has taught me many things, shown me the beauty of research and enthusiastically supported me. I am also grateful to Prof. Christopher H. Schmid for his considerable expertise, kindness and my terrific visiting period at Brown Univeristy. Finally, I wish to extend my special thanks to my colleagues, professors and other people at the Department of Statistical Science, University of Padova, for doing many great things to me and being a wonderful part of my Ph.D journey.

Contents

List of Figures

List of Tables

Introduction

Overview

Meta-analysis is a long-established approach for the quantitative analysis of information from independent studies about the same issue of interest. Traditionally, meta-analysis has been employed in medical and epidemical investigations. Recent attention has involved other disciplines, including economics, ecology, education (see, e.g., Shadish and Lecy, 2015, Gurevitch *et al.*, 2018). A relevant goal in meta-analysis is the evaluation and explanation of heterogeneity across studies, which can be carried out through the inclusion of study-specific information, or covariates, giving rise to the so-called meta regression (Tipton *et al.*, 2019a).

In meta-analysis of the effectiveness of a treatment, differences between studies can be due to studies' design or patients' characteristics. However, not all the sources of heterogeneity can be quantified, as, for example, the severity of illness in patients. An approximation of the severity of illness is given by the control risk, namely, the proportion or rate of events in the control condition. The inclusion of this information in meta-regression gives rise to the so-called control risk regression (Chaimani, 2015, Guolo *et al.*, 2021). The measures of risk in the treated group and in the control group are surrogates of the true unknown values, which are estimated from each study included in the meta-analysis (see, e.g., van Houwelingen *et al.*, 2002). The consequent measurement error problem should be taken into account in the meta-analytic model in order to avoid fallacious inference (Carroll *et al.*, 2006, Yi, 2017). Several solutions have been proposed in the literature to face the measurement error problem in control risk regression (Guolo *et al.*, 2021). A hierarchical modeling approach considers a linear random-effects meta-analytic model for the true unknown risks of outcome for treated and controls and a measurement error model for the observed measures of risk. The normal approximate error model is the typical choice for computational reasons. Likelihood-based inference is typically performed, with or without flexibly accounting for deviations from normality for the random effects (Arends *et al.*, 2000, Ghidey *et al.*, 2007, Guolo, 2013). Other

solutions are based on the corrected score or conditional score approaches (Ghidey *et al.*, 2013) or make use of simulation-based techniques (Guolo, 2014) derived from the measurement error literature.

In the above-mentioned approaches, control risk is the only covariate considered in the meta-regression model mainly for computational convenience. Additional study-specific covariates are usually not accounted for, although they may help explaining between-study heterogeneity. Their inclusion in the control risk regression model, however, is expected to induce substantial practical issues. Most of study-specific covariates are the results of summarizing individual information. Accordingly, their observations are affected by measurement error and by aggregation bias (Simmonds and Higgins, 2007), which might make the relationship between the covariates and risk measure at individual level and study level conflict. Furthermore, the lack of information about within-study covariances between the risk measures and the error-affected covariates makes likelihood-based inference even more problematic.

A typical feature of control risk regression is the linear relationship assumed between the control risk and the treatment risk. Actually, linearity may not hold in general, or it may only describe a small picture of the true and complex relationship between the measures of risk (Boissel *et al.*, 2008). More flexible solutions could be explored, such as, for example, a quadratic relationship (Arends *et al.*, 2000), leading to a U-shaped behavior, accounting for a positive effect and a negative effect of a treatment. Such a choice, however, comes at the cost of computational effort, since the associated likelihood function loses a closed-form expression.

Main contributions of the thesis

This thesis focuses on some open questions in control risk regression, which are summarized below.

Firstly, attention has been paid to the problem of detecting and explaining between-study heterogeneity, by exploiting the information from study-specific covariates. To this aim, likelihood-based solutions have been proposed for inference accounting for the presence of measurement errors. The lack of information about within-study variance/covariance components is overcome by deriving explicit expressions using Taylor expansions based on study-level covariates' subgroup summary information. As an alternative, a more efficient solution based on a pseudo-likelihood solution is developed, under a working independence assumption between the observed error-prone measures.

The methods perform satisfactorily in a series of simulation studies under different specifications for the sample size, the between-study heterogeneity, as well as the underlying risk distribution. The methods are applied to real meta-analyses about the association between COVID-19 and schizophrenia (Pardamean *et al.*, 2022), and the association between COVID-19 and myocardial injury (Sanz-Sánchez *et al.*, 2021).
A second project of the thesis has started during the period spent at the School of Public Health, Brown University, under the supervision of Professor Christopher H. Schmid. The focus is on a more flexible description of the relationship between measures of risk in the treatment group and in the control group. The classical control risk regression model is modified to a U-shaped relationship between the risk measures, in this way allowing to describe treatments which have a positive effect and a negative effect. The price to pay is in terms of computational issues, since the likelihood function loses a closed-form expression, even under an approximate normal measurement error specification. The method is evaluated in a series of simulation studies, involving scenarios of different sample sizes and between-study heterogeneity, absence or presence of the linear/quadratic relationships between the risk measures. The approach is applied to a meta-analysis about the association between Parkinson's disease and diabetes (Lu *et al.*, 2014), and the meta-analysis about the association between COVID-19 and myocardial injury (Sanz-Sánchez *et al.*, 2021).

Chapter 1

Meta-analysis and meta-regression

Meta-analysis is a well-established process to combine, analyze and summarize seemingly conflicting results from different independent studies that address similar research questions or about the same issue of interest (Sutton and Higgins, 2008). Meta-analysis is recognized as the highest-level tool to construct evidence-based practice. It has quickly become popular in plenty of fields (Shadish and Lecy, 2015). Application of meta-analysis has been mostly seen in medical and epidemiology since 1980. However, it has recently appeared in other sciences, including psychology, ecology, economics, genetics, astronomy, sociological and behavioral sciences. See, for example, extensive illustrations of meta-analysis studies in Koricheva *et al.* (2013) and Gurevitch *et al.* (2018).

1.1 Fixed-effects and random-effects models

Suppose that there are n independent studies about the same issue of interest which provide information about an effect size. The term 'effect size' can be thought of as the effect of a treatment in clinical trials or the strength of a relationship between variables. Let Y_i be the estimate of the effect size θ provided by study i, $i = 1, \cdots, n$. Examples include the standardized mean difference, the log odds ratio, the risk difference or correlation. The choice of effect sizes depends on the type of outcome and the goal of analysis. We refer the interested readers to Chapter 3 in Schmid *et al.* (2021) for further details about different outcomes and effect measures. Together with Y_i, an estimate of the associated uncertainty is provided, usually in form of the estimated standard error s_i^2. As an alternative, a 95% confidence interval for θ can be supplied. Therefore, the minimum information from each study is the pair (Y_i, s_i^2).

The literature distinguishes two main approaches for inference for θ. The simplest meta-analytical model is the fixed-effects model, or the common-effect model, which assumes that every Y_i is sampled from a normal distribution with mean θ and variance σ_i^2 (see,

e.g., Stijnen *et al.*, 2021)

$$Y_i = \theta + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \sigma_i^2\right), \tag{1.1}$$

where variance σ_i^2 is usually assumed known and equal to the estimated variance s_i^2. The assumption about σ_i^2 can be reasonable in case the meta-analysis includes large studies, but can be questionable if the included studies have small sample sizes. See, for example, Van Houwelingen *et al.* 1993, Hamza *et al.* 2008, Bellio and Guolo 2016, Papadimitropoulou *et al.* 2019 for proposals accounting for the uncertainty in the estimation of σ_i^2 through s_i^2. Furthermore, s_i^2 is the only source of variability considered in the model. An estimator of θ can be obtained by pooling estimates Y_i from the recruited studies, taking into account the measure of precision of Y_i through s_i^2

$$\hat{\theta}_{FE} = \frac{\sum_i \omega_{i,FE} Y_i}{\sum_i \omega_{i,FE}}, \quad \omega_{i,FE} = \frac{1}{s_i^2},$$

where the subfix 'FE' stands for 'fixed-effects'. This estimator is a weighted mean in which larger studies with lower s_i^2 have larger weights (Rice *et al.*, 2018). Under standard conditions, $\hat{\theta}_{FE}$ is an unbiased and normally distributed estimator of θ. The associated variance is given by $\left(\sum_i \omega_i\right)^{-1}$ (e.g., Viechtbauer, 2005, Guolo and Varin, 2017).
The second meta-analytic model is the random-effects model which has one more source of variability compared to the fixed-effect model. In other words, it imposes a hierarchical or two-stage structure. First, within every study i, the random-effects model assumes that the estimated effect Y_i is drawn from a normal distribution with mean θ_i and variance s_i^2, as follows,

$$Y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, s_i^2\right), \tag{1.2}$$

where θ_i denotes the underlying effect measure in study i which is unobserved. The first-stage model (1.2) is referred to as the within-study model as it describes the variability within every study. The second-stage model, or the between-study model, then considers the between-study availability, assuming that every study-level effect size θ_i is normally distributed with mean θ and variance τ^2 (DerSimonian and Laird, 1986)

$$\theta_i = \theta + \delta_i, \quad \delta_i \sim N\left(0, \tau^2\right). \tag{1.3}$$

Error terms δ_i and ε_i are assumed to be independent (Thompson and Sharp, 1999, Glasziou and Sanders, 2002, Knapp *et al.*, 2006). Marginally, the estimated effect is normally distributed $Y_i \sim N\left(\theta, s_i^2 + \tau^2\right)$ and the parameter vector of interest is $(\theta, \tau^2)^\top$.

See, for example, Schmidt *et al.* (2009), Borenstein *et al.* (2010), Hunter and Schmidt (2000), Schulze (2004), for further discussion about the choice between the fixed-effects model and the random-effects model specification.

1.1.1 Detecting between-study heterogeneity

Heterogeneity across studies is a renowned topic in meta-analysis which receives much attention from the literature (Hardy and Thompson, 1998, Whitehead, 2002, Viechtbauer, 2007b). Testing homogeneity is equivalent to testing $H_0 : \tau^2 = 0$, a problem usually resolved using a classical Q test based on the Cochran's statistic (Cochran, 1954, 1937)

$$Q_{FE} = \sum_{i=1}^{n} \omega_{i,FE} \left(Y_i - \hat{\theta}_{FE} \right)^2 .$$

Under the null hypothesis of homogeneity among studies, test statistic Q_{FE} follows a χ^2_{n-1} distribution. The null hypothesis H_0 is rejected when the observed value of Q_{FE} is larger than $\chi^2_{n-1;1-\alpha}$, where $\chi^2_{n-1;1-\alpha}$ is the α-th quantile of a χ^2_{n-1} distribution. However, small values of Q_{FE} do not mean the presence of heterogeneity. The test also has low power when the number of studies or the study size is small and hence it is not recommended as an instrument to evaluate homogeneity (Hardy and Thompson, 1998, Viechtbauer, 2007a).

Other ways to detect heterogeneity have also been suggested in the literature. For instance, through simulation, Sanchez-Meca and Marín-Martínez (1997) compare a Q test with some Schmidt-Hunter procedures based on their powers and Type I error rates. They show that the Q test adjusts correctly Type I error rate to the nominal significance level, while the Schmidt-Hunter procedures obtain higher power. In another comparison, Viechtbauer (2007b) concludes that the Q test keeps the tightest control of the Type I error rate if compared to the likelihood ratio test, the Wald test and the score test.

An alternative way to detect heterogeneity is to consider confidence intervals for the between-study variance. Viechtbauer (2007a) proposes the Q-profile confidence interval for τ^2 and shows that its coverage probability is larger compared to the Biggerstaff-Tweedie confidence interval, the profile likelihood confidence interval, the Wald-type confidence interval, the Sidik-Jonkman confidence interval, and the bootstrap confidence interval, especially when assumptions about normally distributed effect size estimates and known within-study variances only hold asymptotically. Moreover, when these assumptions are satisfied, the Q-profile confidence interval reaches the nominal coverage

probability.

Some statistics which are independent of the number of studies and the choice effect measures are also proposed in Higgins and Thompson (2002) to measure the impact of heterogeneity on a meta-analysis, namely,

$$H = \sqrt{\frac{Q_{FE}}{n-1}}, \quad R = \frac{se\left(\hat{\theta}_{RE}\right)}{se\left(\hat{\theta}_{FE}\right)}, \quad I^2 = \max\left\{\frac{Q_{FE} - (n-1)}{Q_{FE}} \times 100\%, 0\right\},$$

where $\hat{\theta}_{RE}$ is an estimate of θ in the random-effects model. The most common statistic is I^2, defined as the percentage of total variation due to between-study heterogeneity. Values of I^2 equal to $25\%, 50\%$ and 75% are tentatively considered low, moderate and high, respectively (Higgins *et al.*, 2003).

Since most of the tools discussed in this section are affected when the number of studies is small, it is recommended to always assume heterogeneity across studies and hence model the estimated effect measures Y_i through a random-effect model (e.g., Borenstein *et al.*, 2010).

1.1.2 DerSimonian-Laird estimator

The most famous estimator of $(\theta, \tau^2)^\top$ is proposed in a seminar paper by DerSimonian and Laird (1986). When τ^2 is known, the authors suggest to estimate the effect size by taking a weighted average of Y_i

$$\hat{\theta}_{DL} = \frac{\sum_i \omega_{i,DL} Y_i}{\sum_i \omega_{i,DL}}, \quad \omega_{i,DL} = \frac{1}{\tau^2 + s_i^2}, \tag{1.4}$$

where each weight is the inverse of the sum of a within-study variance and the between-study variance. Under the random-effects model, $\hat{\theta}_{DL}$ is a uniformly minimum variance unbiased estimator of the true effect measure (see, e.g., Langan *et al.*, 2019). Since $\hat{\theta}_{DL}$ has smaller weights compared to $\hat{\theta}_{FE}$ as τ^2 is also included in $\omega_{i,DL}$, the standard error of $\hat{\theta}_{DL}$ is larger than the standard error of $\hat{\theta}_{FE}$ and thus the associated Wald-type confidence intervals are wider.

In practice, the between-study variance τ^2 is unknown. DerSimonian and Laird propose to estimate it using the method of moments

$$\hat{\tau}^2_{DL} = \max\left\{0, \frac{Q_{FE} - (n-1)}{\sum_i \omega_{i,FE} - \sum_i \omega^2_{i,FE} / \sum_i \omega_{i,FE}}\right\}, \tag{1.5}$$

where the truncation is needed to avoid negative values. By replacing τ in formula (1.4) with $\hat{\tau}^2_{DL}$, we obtain the DerSimonian-Laird (DL) estimator of θ

$$\hat{\theta}_{DL} = \frac{\sum_i \omega_{i,DL} Y_i}{\sum_i \omega_{i,DL}}, \tag{1.6}$$

where

$$\omega_{i,DL} = \frac{1}{\hat{\tau}^2_{DL} + s_i^2}.$$

The standard error of this estimator is

$$\widehat{se}\left(\hat{\theta}_{DL}\right) = \frac{1}{\sqrt{\sum_i \omega_{i,DL}}}.$$

According to the central limit theorem, the DL estimator of the true effect size is asymptotically normally distributed. The approach works reasonably well if the number of studies is relatively large (Jackson *et al.*, 2010), and it has a straightforward implementation. However, the uncertainty due to including $\hat{\tau}^2_{DL}$ in formula (1.6) of $\hat{\theta}_{DL}$ is not taken into account. Thus, the method produces confidence intervals for θ which are narrower on average than they should be. The DL estimator of the between-study variance is positively biased because of the truncation (Viechtbauer, 2005). Moreover, when the number of studies is small and the between-study variance is large, $\hat{\tau}^2_{DL}$ often becomes negatively biased (Langan *et al.*, 2019).

Other moment-based estimators are also proposed in the literature. They assume that the estimator of the true effect is a weighted average of Y_i, then equate the associated Q statistic to its expectation and solve for τ^2, and finally plug in the solution to formula (1.4) to obtain an estimator of θ. The general formula of estimators of τ^2 from moment-based approaches is (Langan *et al.*, 2019)

$$\hat{\tau}^2_M = \max\left\{0, \frac{\sum_i \omega_i \left(Y_i - \hat{\theta}\right)^2 - \sum_i \omega_i s_i^2 + \sum_i \omega_i^2 s_i^2 / \sum_i \omega_i}{\sum_i \omega_i - \sum_i \omega_i^2 / \sum_i \omega_i}\right\}.$$

Different choices of weights have been proposed (Paule and Mandel, 1982, Hedges and Olkin, 1985, DerSimonian and Kacker, 2007, Jackson, 2013, Veroniki *et al.*, 2016, Langan *et al.*, 2019). Although the choice of weights' formulas does not affect the estimation of the true effect measure, it has an impact on the associated standard error and confidence intervals (Stijnen *et al.*, 2021). Some moment-based estimators have a closed-form expression, while others need to be computed numerically. Veroniki *et al.* (2016) review all

papers which compare estimation methods for the between-study heterogeneity variance and recommend to use the Paule-Mandel approach for meta-analyses with continuous and dichotomous outcomes. However, the Paule-Mandel estimator of τ^2 is considerably positively biased if the included studies' sizes are much different from each other (Langan *et al.*, 2019).

1.1.3 Hartung-Knapp-Sidik-Jonkman approach

Inference for the true effect size based on asymptotic normality of the DL estimator of θ becomes unreliable when the number of studies is small. It is thus suggested to adjust the standard error of the estimator of the true effect size in order to account for the uncertainty from $\hat{\tau}^2_{DL}$ (Hartung and Knapp, 2001a,b, Knapp and Hartung, 2003, Sidik and Jonkman, 2002)

$$\begin{aligned}\widehat{se}_{HKSJ}\left(\hat{\theta}_{DL}\right) &= \sqrt{\frac{\sum_i \left(Y_i - \hat{\theta}_{DL}\right)^2 / \left(s_i^2 + \hat{\tau}^2_{DL}\right)}{(n-1)\sum_i 1/\left(s_i^2 + \hat{\tau}^2_{DL}\right)}} \times \sqrt{\frac{1}{\sum_i 1/\left(s_i^2 + \hat{\tau}^2_{DL}\right)}} \\ &= \sqrt{\frac{\sum_i \left(Y_i - \hat{\theta}_{DL}\right)^2 / \left(s_i^2 + \hat{\tau}^2_{DL}\right)}{(n-1)\sum_i 1/\left(s_i^2 + \hat{\tau}^2_{DL}\right)}} \times \widehat{se}\left(\hat{\theta}_{DL}\right), \qquad (1.7)\end{aligned}$$

where a truncation towards one can be implemented when the adjusted standard error is smaller than the unadjusted standard error. When $Y_i \sim N\left(\theta, \tau^2 + s_i^2\right)$,

$$\frac{\hat{\theta}_{DL} - \theta}{\widehat{se}_{HKSJ}\left(\hat{\theta}_{DL}\right)} \sim t_{n-1}.$$

Therefore, it is further suggested to perform the inference using a Student t distribution with $n-1$ degrees of freedom in place of the standard normal distribution for statistic $t = \left(\hat{\theta}_{DL} - \theta\right) / \widehat{se}_{HKSJ}\left(\hat{\theta}_{DL}\right)$. This two-step method is referred to as Hartung-Knapp-Sidik-Jonkman (HKSJ) method and is preferred to unadjusted approaches when the number of studies is small. Confidence intervals for the true effect size from this approach are wider than (Wald-type) confidence intervals from the DL approach. Therefore, confidence intervals from HKSJ method also have larger coverage probability compared to the DL approach. In general, the first step, namely, the adjustment of standard errors, can be applied on any estimators of θ and τ^2. However, this method still assumes that every within-study variance is equal to its estimate. Moreover, HKSJ confidence intervals become conservative when the number of studies is small.

1.1.4 Likelihood estimators

The likelihood-based approach is commonly used because of its optimal large-sample properties. Since $Y_i \sim N\left(\theta, \tau^2 + s_i^2\right)$, the associated log-likelihood function for $\left(\theta, \tau^2\right)^\top$ has a closed-form expression, as follows, up to (additive) constants,

$$\ell\left(\theta, \tau^2\right) = -\frac{1}{2}\sum_{i=1}^{n} \log\left(s_i^2 + \tau^2\right) - \frac{1}{2}\sum_{i=1}^{n} \frac{\left(Y_i - \theta\right)^2}{s_i^2 + \tau^2}.$$

This function is also referred to as the approximate log-likelihood function because it is derived from a structure which assumes known within-study variances for the estimated effect sizes (van Houwelingen *et al.*, 2002, Stijnen *et al.*, 2010). Setting the partial derivatives of this function to zero, we obtain the following score equations

$$\hat{\theta}_{MLE} = \frac{\sum_i Y_i / \left(s_i^2 + \hat{\tau}^2_{MLE}\right)}{\sum_i 1 / \left(s_i^2 + \hat{\tau}^2_{MLE}\right)},$$
$$\hat{\tau}^2_{MLE} = \frac{\sum_i \left\{\left(Y_i - \hat{\theta}_{MLE}\right)^2 - s_i^2\right\} / \left(s_i^2 + \hat{\tau}^2_{MLE}\right)^2}{\sum_i 1 / \left(s_i^2 + \hat{\tau}^2_{MLE}\right)^2}.$$

Since the maximum likelihood (ML) estimating equations for θ and τ^2 are connected, the ML estimator can be computed by jointly solving these equations employing simple iterative numerical methods, for example, the Newton-Raphson method (Hardy and Thompson, 1996, Brockwell and Gordon, 2001). The ML estimator of θ is a weighted average of Y_i with weights proportional to $s_i^2 + \tau^2$. Under regularity conditions, the ML estimator of τ^2 is asymptotically unbiased and normally distributed with variance approaching Cramér-Rao lower bound (Veroniki *et al.*, 2016). Therefore, it is common to perform inference based on Wald-type statistics.

Since the ML estimator of τ^2 is known to be biased downwards because of the loss of degree of freedom due to the estimation of θ, a preferable solution is to rely on the restricted likelihood function (Viechtbauer, 2005)

$$\ell_{REML}(\tau^2) = -\frac{1}{2}\sum_{i=1}^{n} \log(s_i^2 + \tau^2) - \frac{1}{2}\sum_{i=1}^{n} \frac{\left\{Y_i - \sum_i Y_i / \left(s_i^2 + \tau^2\right) / \sum_i \left(s_i^2 + \tau^2\right)\right\}^2}{s_i^2 + \tau^2} - \frac{1}{2}\log\left(\sum_{i=1}^{n} \frac{1}{s_i^2 + \tau^2}\right).$$

The restricted log-likelihood function is the marginal log-likelihood function based on residuals $Y_i - \hat{\theta}_{MLE}$. The associated restricted maximum likelihood (REML) estimator can also be computed using iterative numerical methods. Specifically, the REML

estimator of τ^2 is the solution of the following equation

$$\hat{\tau}^2_{REML} = \frac{\sum_i \left[\left\{Y_i - \sum_i Y_i / \left(s_i^2 + \hat{\tau}^2_{REML}\right) / \sum_i 1 / \left(s_i^2 + \hat{\tau}^2_{REML}\right)\right\}^2 - s_i^2\right] / (s_i^2 + \hat{\tau}^2_{REML})^2}{\sum_i 1/(s_i^2 + \hat{\tau}^2_{REML})^2} + \frac{1}{\sum_i 1/(s_i^2 + \hat{\tau}^2_{REML})},$$

which is often approximated with the following equation

$$\hat{\tau}^2_{REML} \approx \frac{\sum_i \left[n/(n-1)\left\{Y_i - \sum_i Y_i / \left(s_i^2 + \hat{\tau}^2_{REML}\right) / \sum_i 1 / \left(s_i^2 + \hat{\tau}^2_{REML}\right)\right\}^2 - s_i^2\right] / (s_i^2 + \hat{\tau}^2_{REML})^2}{\sum_i 1/(s_i^2 + \hat{\tau}^2_{REML})^2}.$$

This approximation becomes exact when the included studies have the same within-study variances, i.e., $s_1^2 = \cdots = s_n^2$ (Viechtbauer, 2007a). The REML estimator of θ is computed using weighted average formula (1.4) with $\tau^2 = \hat{\tau}^2_{REML}$. This method is preferred to the ML approach since it results in nearly unbiased estimators for τ^2 (Stijnen *et al.*, 2021). In Veroniki *et al.* (2016) and Langan *et al.* (2019), it is recommended over many methods for estimating the between-study variance.

The likelihood-based inference for θ can be performed using a Wald-type confidence interval. However, this interval is symmetric and not invariant with respect to model reparameterization. These disadvantages can be avoided by adopting the profile likelihood approach. Let $\hat{\tau}^2_\theta$ denote the constrained maximum likelihood estimator of τ^2 for a fixed value of θ and let $\ell_P(\theta) = \ell\left(\theta, \hat{\tau}^2_\theta\right)$ denote the profile log-likelihood function for θ, respectively. The signed profile log-likelihood ratio is

$$r_P(\theta) = \operatorname{sgn}\left(\hat{\theta}_{MLE} - \theta\right)\sqrt{2\left\{\ell_P\left(\hat{\theta}_{MLE}\right) - \ell_P(\theta)\right\}},$$

where $\ell_P\left(\hat{\theta}_{MLE}\right) = \ell\left(\hat{\theta}_{MLE}, \hat{\tau}^2_{MLE}\right)$. Under regularity conditions, $r_P(\theta)$ is asymptotically normally distributed up to an error of order $O\left(n^{-1/2}\right)$ (Severini, 2000). The associated $100(1-\alpha)\%$ confidence interval for θ is $\{\theta : z_{\alpha/2} \leq r_P(\theta) \leq z_{1-\alpha/2}\}$, where z_α is the α-th quantile of the standard normal distribution. A hypothesis test for θ at $100\alpha\%$ significance level is merely based on the comparison of the value of $r_P(\theta)$ under the null hypothesis with $z_{\alpha/2}$ and $z_{1-\alpha/2}$.

When the number of studies is small, likelihood-based inference relying on first-order approximations can be prone to misleading results. Guolo (2012) suggests to perform a second-order adjustment to the signed profile log-likelihood ratio statistic $r_P(\theta)$ using

Skovgaard's statistic (Skovgaard, 1996)

$$r_P^{\star}(\theta) = r_P(\theta) + \frac{1}{r_P(\theta)} \log \frac{u(\theta)}{r_P(\theta)},$$

where $u(\theta)$ is a function of the observed information evaluated at the unconstrained ML estimate and the sample space derivatives of likelihood quantities with respect to the ML estimates. The refined statistic converges to a normally distributed variable with better accuracy $O\left(n^{-1}\right)$ and maintains a computationally feasible form. Through simulation, Guolo (2012) shows that Skovgaard's statistic results in better confidence intervals and rejection rates compared to the first-order counterpart when the number of studies is small to moderate. However, Skovgaard's statistic might suffer from some computational difficulties and numerical instabilities.

Alternatively, Huizenga *et al.* (2011) propose to improve the performance of the signed profile log-likelihood ratio $r_P^2(\theta)$ by applying Bartlett's correction

$$(1 + A)^{-1} r_P^2(\theta) \sim \chi_1^2,$$

where $(1 + A)^{-1}$ is the Bartlett's correction factor. The associated $100(1 - \alpha)\%$ confidence interval for the true effect size is $\left\{\theta : (1 + A)^{-1} r_P^2(\theta) < \chi_{1;1-\alpha}^2\right\}$, where $\chi_{1;1-\alpha}^2$ is the α-th quantile of the chi square distribution with one degree of freedom χ_1^2. The test based on the corrected ratio has Type I error rate closer to the nominal level compared to the Wald-type test, especially when there are a few studies.

As an alternative to the profile likelihood approach, Bellio and Guolo (2016) propose the use of the integrated likelihood approach for inference for the true effect size for continuous outcomes when the number of studies is small. The integrated likelihood method eliminates the nuisance parameters given by the between study variance and within-study variances through integration with respect to a weight function. The integrated log-likelihood is (Severini, 2000)

$$\ell_I(\theta) = \log \left[\int_0^{\infty} \left\{ \prod_{i=1}^{n} g_i(\theta, \zeta) \right\} \pi(\zeta) d\zeta \right],$$

where

$$g_i(\theta, \zeta) = \int_0^{\infty} L\left(\theta, \zeta, \sigma_i\right) \pi(\sigma_i) d\sigma_i, \quad \zeta = \tau^2 - \left(\hat{\theta}_{MLE} - \theta\right)^2,$$

and $\pi(\zeta)$ denotes a weight function specified in Bellio and Guolo (2016). Inference for θ can be performed based on the signed integrated log-likelihood ratio statistic

$$r_I(\theta) = \text{sgn}\left(\tilde{\theta} - \theta\right)\sqrt{2\left\{\ell_I\left(\tilde{\theta}\right\} - \ell_I(\theta)\right)},$$

where $\tilde{\theta} = \arg\max_{\theta}\{\ell_I(\theta)\}$. The approach is shown to provide confidence intervals with coverage probability closer to the nominal level compared to the profile likelihood method. It also avoids numerical issues related to the estimation of heterogeneity which might affect the inference for the true effect.

1.1.5 Nonparametric hypothesis tests for the true effect

Normality assumption for random effects' distribution in meta-analysis is often criticized in the literature (e.g., van Houwelingen *et al.*, 2002, Guolo, 2012), despite its computational convenience. Nonparametric solutions avoid distributional assumptions. Follmann and Proschan (1999) propose a test for the true effect size θ. Under the assumption of symmetric estimated effect sizes and the null hypothesis of no effect, this test permutes the sign of Y_i N times. This approach is equivalent to randomly switching the label of the treatment group and the control group in each study. In every permutation, the test recomputes the value of the chosen test statistic, for example, a z statistic or a t statistic. The reference distribution is the empirical distribution of the N obtained values and the p-value is the proportion of values larger than the value evaluated from the original data. The test controls Type I error rate for typical meta-analyses scenarios (Follmann and Proschan, 1999). It is also robust to model misspecification. However, this test is computational expensive, so it is often suggested to consider a subset of permutations rather than a full set. When the number of studies is very small, it might be impossible to obtain conventional significance levels (Viechtbauer, 2010). Furthermore, this test can also have low power compared to parametric tests.

1.1.6 Bayesian inference

Bayesian inference can be performed on study-level effects θ_i, the true effect size θ and the between-study variance τ^2. In other words, the between-study model $\theta_i \sim N(\theta, \tau^2)$ shows a prior distribution for θ_i. Moreover, study-level effect sizes are exchangeable since their joint distribution does not depend on their order. Their exchangeability reflects a degree of prior ignorance where their magnitudes cannot be differentiated (Higgins *et al.*, 2009), which allows to estimate and infer a study-level effect measure

using information from the other studies (Schmid *et al.*, 2021).

Let $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^\top$ denote a $n \times 1$ vector of estimated effect sizes. If the true effect size is normally distributed $\theta \sim N\left(\theta_0, \sigma_\theta^2\right)$, regardless of the choice of priors for the between-study variance, the conditional posterior distributions of $\theta|\tau^2, \boldsymbol{Y}$ and $\theta_i|\theta, \tau^2, \boldsymbol{Y}$ have closed-form expression, as follows,

$$\theta|\tau^2, \boldsymbol{Y} \sim N\left(\theta_1, \sigma_1^2\right),$$
$$\theta_i|\theta, \tau^2, \boldsymbol{Y} \sim N\left(\hat{\theta}_i, \frac{1}{1/s_i^2 + 1/\tau^2}\right),$$

where

$$\theta_1 = \frac{\hat{\theta}_{DL} \sum_i 1/(\tau^2 + s_i^2) + \theta_0/\sigma_\theta^2}{\sum_i 1/(\tau^2 + s_i^2) + 1/\sigma_\theta^2}, \quad \sigma_1^2 = \frac{1}{1/\sigma_\theta^2 + \sum_i 1/(\tau^2 + s_i^2)}, \quad \hat{\theta}_i = \frac{\tau^2 Y_i + s_i^2 \theta}{\tau^2 + s_i^2}.$$

Therefore, Gelman *et al.* (2013) and Schmid *et al.* (2021) describe a simple algorithm to sample the parameters of interest which starts by numerically generating τ^2 from its posterior distribution $f\left(\tau^2|\boldsymbol{Y}\right)$ and then generate θ from $N\left(\theta_1, \sigma_1^2\right)$ with the obtained value of τ^2. Next, this algorithm generates θ_i from $N\left[\hat{\theta}_i, 1/\left\{\left(s_i^2\right)^{-1} + \left(\tau^2\right)^{-1}\right\}^{-1}\right]$ with the updated values of θ and τ^2. Finally, Y_i is generated from the random-effects model and the algorithm returns to the first step, until convergence. The computation of the full marginal posteriors of θ and τ^2 requires numerical methods such as Markov chain Monte Carlo (MCMC) method because of their complex analytic formulas.

A Bayesian method is complete with the choice of prior distributions for θ and τ^2, which are usually assumed to be independent. While a common choice for the true effect size is a normal distribution $N\left(\theta_0, \sigma_\theta^2\right)$ with a quite large value of variance σ_θ^2, the literature suggests to adopt informative priors for τ^2 which can be constructed based on the past meta-analyses. For instance, in case of dichotomous outcomes, Turner *et al.* (2012, 2015) derive log-normal priors which account for the type of outcome and the type of intervention. In case of continuous outcomes, Rhodes *et al.* (2015) propose log-t priors which also account for studies' characteristics. These data-based priors substantially reduce the uncertainty of the estimators of θ and τ^2.

Although Bayesian methods naturally allow for full uncertainty, especially in predicting study-level effect sizes, they may suffer from computational intensity and sensitivity to priory judgement.

1.1.7 Flexible random-effect distributions

Normality assumption in the between-study model is a strong assumption which might affect the conclusions of inference (Higgins *et al.*, 2009). It is thus suggested to assume flexible parametric distributions for study-level effect sizes. For instance, Smith *et al.* (1995) and Lee and Thompson (2008) suggest a t distribution which gives more weight to outlying studies compared to a normal distribution. Lee and Thompson (2008) propose to consider skewed distributions to allow for potential skewing. Böhning (2000) uses mixture distributions, in order to account for studies belonging to unknown groupings.

Relatively complex models, such as non-parametric likelihood distributions and Bayesian semi-parametric distributions, are suitable choices for the random effect distribution when there is a large number of studies, since they let the observed data determine the shape of the random effect distribution (Higgins *et al.*, 2009). The non-parametric likelihood model results in a discrete distribution that is based on a finite number of mass points (Laird, 1978, Böhning, 2005). Therefore, non-parametric likelihood distributions are able to detect and incorporate outliers, but they are unstable (Van Houwelingen *et al.*, 1993). The Bayesian semi-parametric model is based on a Dirichlet process prior (Burr *et al.*, 2003, Burr and Doss, 2005, Ohlssen *et al.*, 2007). Similar to non-parametric likelihood models, Bayesian semi-parametric models can result in predictive distributions which have unconventional shapes and strongly depend on studies at hand (Higgins *et al.*, 2009).

1.2 Meta-regression

Besides the use of the between-study variance, the variability across studies can be explained by study-level characteristics which include the methodological features of studies, the descriptors of study context, the descriptors of participants, the characteristics of experiment interventions and exposures being evaluated in included studies, and the aspects of outcome being measured (Tipton *et al.*, 2019b,a). These characteristics are encoded with study-specific covariates.

It is first suggested in Glass and Smith (1979) to fit the ordinary least squares regression to meta-analysis with covariates. However, such a solution does not account for heteroscedasticity in the estimated effect sizes, i.e., $var\left(Y_i\right) = \tau^2 + s_i^2$ changes across studies. As an alternative, Raudenbush and Bryk (1985) extend the random-effects model by keeping the within-study model (1.2) and assuming the linear relationship between covariates and the study-level effects instead of the estimated effects. This

solution gives rise to the so-called meta-regression model. Besides subgroup analysis, meta-regression is an instrument to quantify the contribution of study-level characteristics to the heterogeneity across studies. Let $X_1, \cdots, X_k$ denote k study-level covariates. A meta-regression model is defined as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \delta_i + \varepsilon_i, \quad \delta_i \sim N\left(0, \tau^2\right), \quad \varepsilon_i \sim N\left(0, s_i^2\right), \tag{1.8}$$

where τ^2 denotes the residual variance which shows the heterogeneity across studies that is unexplained by the study-level covariates.
Consider a meta-regression model of n studies. Let $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^\top$ denote an $n \times 1$ vector of estimated effect sizes and $\boldsymbol{X}$ denote an $n \times (k+1)$ design matrix of full column rank, respectively. The meta-regression model (1.8) can be rewritten in the matrix form

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\delta} \sim N\left(0, \tau^2 I_n\right), \quad \boldsymbol{\varepsilon} \sim N\left\{0, V = \operatorname{diag}\left(s_i^2\right)\right\},$$

where $\boldsymbol{\beta} = \left(\beta_0, \beta_1, \cdots, \beta_k\right)^\top$ is a $(k+1) \times 1$ vector of regression coefficients, $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_n)^\top$ is an $n \times 1$ vector of residuals, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)^\top$ is a $n \times 1$ vector of within-study errors, and V denotes the diagonal matrix of within-study variances. The parameter of interest is thus $\theta = \left(\beta_0, \beta_1, \cdots, \beta_k, \tau^2\right)^\top$.

1.2.1 Estimation

A moment-based approach to estimate θ starts by estimating regression coefficients using the weighted least squares method (Raudenbush, 2009),

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^\top W \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top W \boldsymbol{Y}, \tag{1.9}$$

where $W = \operatorname{diag}(\omega_i)$ is an $n \times n$ diagonal matrix of weights. Then the method sets the associated weighted residual sum of squares to its expectation and solves for τ^2. The estimated residual variance $\hat{\tau}_M^2$ is then plugged in the following formula for estimating regression coefficients,

$$\hat{\boldsymbol{\beta}}_M = \left(\boldsymbol{X}^\top W_M \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top W_M \boldsymbol{Y}, \quad \widehat{var}\left(\hat{\boldsymbol{\beta}}_M\right) = \left(\boldsymbol{X}^\top W_M \boldsymbol{X}\right)^{-1}, \tag{1.10}$$

where $W_M = \operatorname{diag}\left\{\left(s_i^2 + \hat{\tau}_M^2\right)^{-1}\right\}$ is an $n \times n$ diagonal matrix of estimated weights. A truncation toward zero is used when obtaining negative estimates of the residual variance. The choice of weight matrices in the first step is different across moment-based

approaches. See, e.g., Raudenbush (1994), Sidik and Jonkman (2005a,b), Raudenbush (2009), López-López *et al.* (2014), Viechtbauer *et al.* (2015), for different moment-based approaches. Moment-based estimators of regression coefficients are asymptotically normally distributed. If the residual variance is known, weighted least squares estimators of the regression coefficients are unbiased, i.e., $E\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta}$. The choice of weight matrices W does not influence the estimation of regression coefficients but the associated standard error and thus inference. However, moment-based inference for regression coefficients might be unreliable since the uncertainty of the estimated residual variance is not taken into account when estimating regression coefficients in the last step.
It is thus preferable to use the maximum likelihood approach to account for the uncertainty in the estimator of the residual variance. Under normality assumption of the estimated effects,

$$\boldsymbol{Y} \sim N_n\left(\boldsymbol{X}\boldsymbol{\beta}, V + \tau^2 I\right).$$

Therefore, the log-likelihood function for θ is, up to constants (Raudenbush, 2009, Viechtbauer *et al.*, 2015)

$$\ell\left(\theta\right) \propto -\frac{1}{2}\log\left|V + \tau^2 I\right| - \frac{1}{2}\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\right)^\top W\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\right).$$

Under regularity conditions, the maximum likelihood (ML) estimator of θ is asymptotically normally distributed with mean θ and covariance matrix $-J^{-1}$, where J is the observed Fisher information. Nevertheless, the ML estimator of the residual variance is negatively biased due to the loss of degree of freedom when estimating regression coefficients. This problem can be resolved by using the restricted maximum likelihood (REML) method, which removes $\boldsymbol{\beta}$ from $\ell\left(\theta\right)$ through integration (Raudenbush, 2009, Viechtbauer *et al.*, 2015)

$$\ell_{REML}\left(\tau^2\right) \propto \frac{1}{2}\log\left|\boldsymbol{X}^\top\boldsymbol{X}\right| - \frac{1}{2}\log\left|V + \tau^2 I\right| - \frac{1}{2}\log\left|\boldsymbol{X}^\top W\boldsymbol{X}\right| - \frac{1}{2}\boldsymbol{Y}^\top M\boldsymbol{Y},$$

where

$$M = W - W\boldsymbol{X}\left(\boldsymbol{X}^\top W\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top W.$$

While the REML estimator of regression coefficients has the same asymptotic behavior as the ML estimator, the REML estimator of the residual variance is nearly unbiased. The ML estimator and REML estimator are computed by equating the associated score functions to zero and solving the resulting equations with iterative numerical algorithms,

for example, the Newton-Raphson algorithm. Alternatively, Raudenbush (2009), López-López *et al.* (2014), Viechtbauer *et al.* (2015) describe an algorithm to compute the likelihood-based estimators which starts by choosing a value of $\hat{\tau}^2$ from non-iterative approaches such as moment-based ones. Then it updates the weight matrix in formula (1.9) via the following formula

$$\omega_i = \frac{1}{s_i^2 + \hat{\tau}^2_{\text{old}}},$$

and updates the estimator of the residual variance by adding an amount of Δ to the previous estimate $\hat{\tau}^2_{\text{old}}$. This step is repeated until convergence is obtained. Finally, a likelihood-based estimator of regression coefficients follows formula (1.9). The formula of the updated amount Δ is

$$\Delta_{ML} = \frac{\boldsymbol{Y}^\top MM\boldsymbol{Y} - \text{tr}(W)}{\text{tr}(WW)},$$

for the maximum likelihood approach and

$$\Delta_{REML} = \frac{\boldsymbol{Y}^\top MM\boldsymbol{Y} - \text{tr}(M)}{\text{tr}(MM)}$$

for the restricted maximum likelihood approach.
It is also possible to perform Bayesian inference for θ with, e.g., MCMC approaches or the Gibbs sampling approach (Sutton and Abrams, 2001, Spiegelhalter *et al.*, 2003, Higgins *et al.*, 2009). Specifically, it is common to assume vague priors such as normal distributions with large variance for regression coefficients. For the residual variance, it is possible to adopt priors which are similar to priors of the between-study variance in a random-effects model. Likelihood-based approaches and non-informative Bayesian approaches provide similar results when there are many studies, while credible intervals for regression coefficents become wider than the corresponding confidence intervals when the number of studies decreases (Schmid *et al.*, 2021).

1.2.2 Hypothesis testing and confidence intervals

Wald-type inference for $\beta_j, j = 1, \cdots, k$, can be carried out straightforwardly starting from the moment-based estimators. However, since moment-based approaches do not account for the uncertainty in the standard errors of the estimators of regression coefficients due to the estimator of the residual variance, the empirical coverage probabilities of Wald-type confidence intervals are smaller than the nominal coverage probability and

the associated hypothesis tests do not adequately control Type I error (e.g., Tipton *et al.*, 2019a).

As an alternative, when data are in form of 2×2 tables, Berkey *et al.* (1995) suggest to approximate the sampling distribution of the centering estimator $\left(\hat{\beta}_j - \beta_j\right) / \widehat{se}\left(\hat{\beta}_j\right)$ as a Student t-distribution with $n - k - 3$ degrees of freedom since the t distribution penalizes meta-analyses with a few studies and/or many covariates (Tipton *et al.*, 2019a). To account for the uncertainty due to the estimator of the residual variance, Knapp and Hartung (2003) (KH) propose to adjust the estimator of the standard error of the estimator of β_j, as follows,

$$\widehat{se}_{KH}\left(\hat{\beta}_j\right) = \sqrt{\frac{1}{n-k-1}\sum_{i=1}^{n}\frac{\left\{Y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}\right)\right\}^2}{s_i^2 + \hat{\tau}^2}} \times \widehat{se}\left(\hat{\beta}_j\right),$$

where a truncation towards one can be used when the adjusted standard error is smaller than the unadjusted standard error. They also define the t statistic, as follows,

$$t = \frac{\hat{\beta}_j - \beta_j}{\widehat{se}_{KH}\left(\hat{\beta}_j\right)},$$

and show that it follows a Student t-distribution with $n - k - 1$ degrees of freedom. Hypothesis tests based on the Knapp-Hartung approach maintain their nominal Type I error rates across a wide range of conditions (Higgins and Thompson, 2004, Viechtbauer *et al.*, 2015). However, this method can be over-conservative due to the truncation when the number of studies is small.

Since Wald-type confidence intervals for the regression coefficient β_j based on the ML estimator are symmetric and not invariant to model reparameterization, a signed profile likelihood ratio test statistic can be used (Huizenga *et al.*, 2011)

$$r_P\left(\beta_j\right) = \operatorname{sgn}\left(\hat{\beta}_{j,MLE} - \beta_j\right)\sqrt{2\left\{\ell_P\left(\hat{\beta}_{j,MLE}\right) - \ell_P\left(\beta_j\right)\right\}},$$

where $\ell_P\left(\beta_j\right)$ is the profile log-likelihood function for β_j. The test statistic $r_P\left(\beta_j\right)$ converges to a normal distribution with an error of order $O\left(n^{-1/2}\right)$. Several adjusted versions of the test statistic are also proposed to improve its accuracy and hence the associated Type I error rate. For instance, Huizenga *et al.* (2011) apply Bartlett's correction to the profile likelihood ratio test statistic $r_P^2\left(\beta_j\right)$ to reduce the deviation of the likelihood ratio test from a chi square distribution χ_1^2, especially when the number of studies is small. Or Guolo (2012) proposes a second-order profile likelihood ratio test

statistic based on the Skovgaard's statistic (Skovgaard, 1996)

$$r_P^{\star}(\beta_j) = r_P(\beta_j) + \frac{1}{r_P(\beta_j)} \log \frac{u(\beta_j)}{r_P(\beta_j)},$$

where $u(\beta_j)$ is the function of the observed information evaluated at the unconstrained ML estimate of θ and the sample space derivatives of likelihood quantities with respect to the ML estimates. This second-order profile likelihood ratio test statistic is asymptotically normally distributed up to an error of order $O\left(n^{-1}\right)$. Therefore, this test statistic has empirical rejection rate and empirical coverage probability very close to the nomial level. The profile likelihood ratio test with the Bartlett's correction and the second-order profile likelihood ratio test outperform Wald-type tests with and without the KH approach, especially when the number of studies is small (Huizenga *et al.*, 2011, Guolo, 2012).

To avoid the distributional assumptions of random effects and within-study errors, non-parametric tests are suggested, especially when the number of studies is small. A non-parametric test proposed in Higgins and Thompson (2004) starts by computing a t statistic and then permutes the index of link pairs $\left(Y_i, s_i^2\right)^{\top}$ several times. In each permutation, the test recomputes the t statistic. The p-value of this test is the proportion of permutations whose values of the t statistic are larger or equal to the value in the first step. This permutation test adequately controls Type I error rate compared to the other tests in a wide range of scenarios (Higgins and Thompson, 2004, Viechtbauer *et al.*, 2015). Huizenga *et al.* (2011) evaluate the significance of regression coefficient β_j using a resampling test which starts by computing the statistic $\left(\hat{\beta}_j - \beta_j\right) \Big/ \widehat{se}\left(\hat{\beta}_j\right)$ based on the full model and computing the residuals when fitting the meta-regression model without covariate X_j. Then this test resamples n times without replacement from the obtained residuals, adds the resampled residuals to the models for the included studies, reorders the elements of V accordingly, and recomputes the statistic $\left(\hat{\beta}_j - \beta_j\right) \Big/ \widehat{se}\left(\hat{\beta}_j\right)$ based on the full model. This step is repeated several times and the p-value of the test is the proportion of values of the resampled statistic which exceed the value from the first step. Through simulation, Huizenga *et al.* (2011) shows that their resampling test has slightly more accurate Type I error rate and slightly less power compared to the profile likelihood ratio test with the Bartlett's correction. However, this resampling test and the permutation test in Higgins and Thompson (2004) might not reach some conventional significance levels and suffer from heavy computational expense.

1.2.3 Type I and Type II errors

Inference in meta-regression models with very large number of covariates might yield to some issues with Type I error. As a solution, Hunter and Schmidt (2015) propose to focus on covariates to avoid inflated Type I error. The choice of covariates is important, especially when the number of studies is small to moderate (Raudenbush, 1994) because choosing covariates without a strong reason may lead to spurious findings as a result of chance, even if the number of studies is large (Schmidt and Hunter, 2015). Another way to control Type I error rate is to use adjustment methods, for example, the KH approach, the Bartlett's correction, and the second-order test statistic. When testing multiple regression coefficients and reporting statistically significant ones, (Thompson and Higgins, 2002) propose to prespecify covariates and use multiple-comparisons corrections to avoid very high family-wise error rates. Alternatively, Type I errors can be controlled by concentrating on the magnitude of the estimates of regression coefficients and the practical importance of values entertained by confidence intervals (Sterne and Smith, 2001).

Risk of low power to detect relationship is also an issue in meta-regression. The failure to find associations between the estimated effect size and covariates is due to the insufficient number of studies or limited variability in the values of covariates. Through simulation, Schmid *et al.* (2004) show that a meta-regression model is effective for detecting associations between effect sizes and covariates if there are at least ten studies. The approaches in Subsection 1.2.2 have power under 80% when being applied on meta-analyses with at most 20 studies (Huizenga *et al.*, 2011, Viechtbauer *et al.*, 2015). Through simulation, Berkey *et al.* (1995) show that the power of a test for a regression coefficient associated to a single continuous covariate with moderate effect can be quite low when the number of studies is small. Some analytic methods are derived in Hedges and Pigott (2004) to calculate the power of a test for coefficients, which show that the power is often quite low. Furthermore, Hedges and Pigott (2004) propose to perform power analyses in the review process and avoid performing hypothesis tests for regression coefficients when the power is low.

1.2.4 Aggregation bias

Meta-regression models are characterized by the presence of study-level covariates which are summarized information of individual characteristics provided by each study included in the meta-analysis. The use of aggregate information can give rise to aggregation bias, or ecological fallacy, which is a term referred to situations where the

relationship between the effect size and covariates at the individual level is different from the relationship at the study level (Simmonds and Higgins, 2007, Baker *et al.*, 2009). Aggregation bias commonly occurs when effect sizes are reasonably homogeneous across studies, a condition resulting in no relationship between the effect size and the covariates, although there might exist an association between them at the individual level. Ecological fallacy may also occur when the values of a covariate are similar at the study level (Berlin *et al.*, 2002). Besides, confounders can result in an association found at the study level that does not exist at the individual level. Aggregation bias might also occur when the values of a covariate at the individual level largely fluctuates since the summary of these values at each study does not accurately represent their sample in that study (Baker *et al.*, 2009). To reduce this bias, when the covariate of interest is the prevalence of a category, Schmid *et al.* (2004) suggest not to extrapolate the effect of a change in the prevalence at the population level to the individual level except in a vague directional sense.

Chapter 2

Classical control risk regression

Meta-analysis is commonly used in studies aimed at evaluating the effect of a treatment based on the comparison of a treatment group with a placebo or control group. The contribution of study-level characteristics to the heterogeneity across studies can not be quantified by meta-regression if these characteristics are not measured or available. In this case, the underlying or baseline outcome event rate or mean for patients in the control group provides a surrogate measure that is always available from studies included in the meta-analysis. The inclusion of the baseline risk as a covariate in a meta-regression model gives rise to the so-called control risk regression model. See, e.g., Schmid *et al.* (1998), Arends *et al.* (2000), van Houwelingen *et al.* (2002), Guolo *et al.* (2021).

2.1 Model for treatment risk given control risk

Consider a meta-analysis of n independent studies about the effectiveness of a treatment. Let η_i and ξ_i denote risk measures in the treatment group and in the control group of study i, $i = 1, \cdots, n$, respectively. Let $\eta_i^\star = \eta_i - \xi_i$ denote the treatment effect. The classical control risk regression model assumes that the measures of risk are related by a linear model (Arends *et al.*, 2000, van Houwelingen *et al.*, 2002)

$$\eta_i = \beta_0 + \beta_1 \xi_i + \varepsilon_i, \quad \varepsilon_i \sim N\left(0, \tau^2\right), \tag{2.1}$$

where τ^2 is the variance of the treatment risk across studies that is unexplained by the control risk. The events under consideration are negative. If the intercept is zero, a value of β_1 smaller than one corresponds to the effectiveness of the treatment, while a value of β_1 larger than one corresponds to the harmfulness of the treatment. There is no relationship between the treatment effect and the control risk when β_1 is equal to

one, so it is more interesting to see values of β_1 different from one. The control risk regression model can also be defined as a linear model of the treatment effect $\eta_i^\star$ and the baseline risk ξ_i (see, e.g., Brand and Kragt, 1992, McIntosh, 1996, Schmid *et al.*, 1998)

$$\eta_i^\star = \beta_0^\star + \beta_1^\star (\xi_i - \mu_\xi) + \varepsilon_i^\star, \quad \varepsilon_i^\star \sim N\left(0, \tau^{\star 2}\right). \tag{2.2}$$

Model (2.2) is actually a reparameterization of model 2.1, where $\beta_0^\star$ denotes the average treatment effect when the control risk is equal to its expected value μ_ξ. Specifically, model (2.1) can be re-expressed as model (2.2) by simply subtracting ξ_i from two sides of model (2.1) and centering this variable. A negative value of $\beta_1^\star$ indicates that $\eta_i^\star$ decreases with ξ_i, i.e., the treatment becomes more effective when the disease is more serve in population under control condition. Similarly, a positive value of $\beta_1^\star$ corresponds to the harmfulness of the treatment. A value of $\beta_1^\star$ equal to zero shows no relationship between the treatment effect and the baseline risk. While model (2.2) is more computationally efficient and intuitive than the control risk regression model with treatment risk, it has some drawbacks resulted from the negative correlation between the estimated measures of treatment effect and the control risk since the former depends on the latter (van Houwelingen *et al.*, 2002). Therefore, its application is not recommended and we will mainly consider the model with the treatment risk in this thesis.
The linear relationship between the control risk and the treatment risk is mainly based on empirical consideration and computational convenience. However, through simulation, Boissel *et al.* (2008) discover that the linear relationship only holds in a short range of frequency of event and hence shows an incomplete picture of the true relationship between the measures of risk, due to the limited number of studies. Therefore, flexible relationships between the true risk measures have been discussed or derived in the literature. Examples include the quadratic model Arends *et al.* (2000) and a model resulting a U-shaped relationship between the absolute risk difference and the control risk Wang *et al.* (2009). Further details can be found in Chapter 4.

2.1.1 L'Abbé plot

The true unobserved measures of risk in model (2.1) are estimated with the proportion or the rate of diseased participants in every group and study. The observed or estimated risk measures serve as a surrogate for the true ones. A graphical way to evaluate the treatment effect is via the L'Abbé plot (L'Abbé *et al.*, 1987) which is a scatter plot of the observed control risk measure (x-axis) and the observed treatment risk measure (y-axis). Every data point in a L'Abbé plot is shown in form of a circle

With MI		Without MI	
Events	Total	Events	Total
68	94	15	109
13	21	28	158
50	123	18	209
12	23	16	105
31	52	12	135
14	24	8	30
46	112	26	112
504	914	302	1906
48	89	15	35
51	133	11	538
121	170	65	989
0	10	1	125
3	16	0	85
23	50	1	95

TABLE 2.1: Myocardial injury dataset (Sanz-Sánchez *et al.*, 2021).

with size proportional to the amount of information from the associated study. For example, the circle's size can be the inverse of the standard error of the associated observed treatment effect or the inverse of the associated study size. Thus, large circles correspond to studies with large sample size and precise estimate of the treatment effect. Originally, the L'Abbé plot is suggested to show the variation of the treatment effects (L'Abbé *et al.*, 1987). When the risk difference is chosen as the treatment effect, the heterogeneity is present if data points are not close to any lines which are parallel to the identity line, i.e., $\eta \neq \text{constant} + \xi$. When the treatment effect is the risk ratio, the heterogeneity occurs if data points are not close to any lines which pass through the origin, i.e., $\eta \neq \text{constant} \times \xi$. While treatment effects are homogeneous regardless of choices of treatment effect if data points are close to the identity line, different choices of treatment effect result in contrary conclusions on the heterogeneity. Therefore, it is recommended to choose the same risk measures for the L'Abbé plot and fitting a control risk regression model. Moreover, L'Abbé plot can be a graph of the treatment effect against the control risk (Sharp *et al.*, 1996) or a graph of the treatment effect against the average of the control risk measure and the treatment risk measure (Bland and Altman, 1986), which will not be discussed in this thesis.

Table 2.1 shows a meta-analysis of 14 studies in Sanz-Sánchez *et al.* (2021) which evaluates the association between myocardial injury (MI) and COVID-19. Every study reports the total number of COVID-19 patients and the number of deaths due to all causes in the group with MI and the group without MI.

As an example of L'Abbé plot, Figure 2.1 shows a L'Abbé plot for the myocardial

injury dataset (Sanz-Sánchez *et al.*, 2021), where the log odds of all-cause mortality is the measure of risk and the log odds ratio is the treatment effect. The identity line (the thick solid line) is also referred to as the no-effect line. All the circles do not scatter around any lines which are parallel to the identity line, indicating the presence of heterogeneity of treatment effects. Most of the circles are under a line whose slope is smaller than one. The MI dataset is of of examples that the thesis focuses on and will be examined in detail later in Section 3.4.

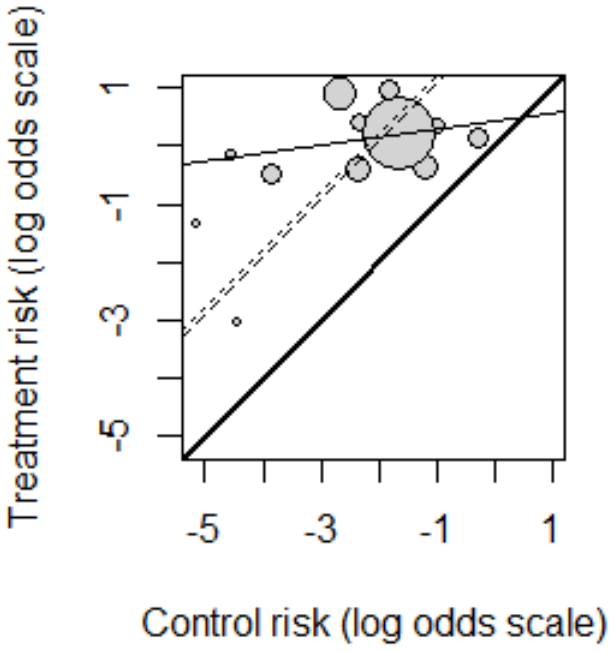

FIGURE 2.1: L'Abbé plot for the myocardial injury dataset (Sanz-Sánchez *et al.*, 2021). The thick solid line is the identity line. The thin solid line is the graph of the linear model fitted by the weighted least squares approach.

Although L'Abbé plot is a useful graphical way to detect heterogeneity, this plot should not be used to assess the treatment effect since it only shows the measures of risk estimated based on samples in the included studies.

2.1.2 Weighted least squares method

Brand and Kragt (1992) propose to fit a control risk regression model using the weighted least squares (WLS) approach. This method estimates $(\beta_0, \beta_1)^\top$ in model (2.1) with the inverses of the variances of the observed measures of treatment risk as weights. Inference for the regression coefficients is based on the Wald-type test and confidence

intervals. The thin solid line in Figure 2.1 is the graph of the linear model fitted with the WLS approach. The slope of this line is smaller than one. The x-coordinate of the intersection of the line and the identity line is referred to as the break-even point, which is defined as a level beyond that the treatment effect is negative and below that the treatment effect is positive. In the MI dataset, the break-even point is between zero and one.

The WLS approach has two major disadvantages despite of simplicity. First, since the measure of treatment effect is a function of the risk measures, errors in estimating the treatment effect and the control risk are negatively correlated. Therefore, if the WLS is applied on model (2.2), negative correlation is not accounted for and hence it yields a negative bias in the estimator of $\beta_1^\star$. A suggested solution is to apply the WLS approach on models where errors in the dependent variable and the independent variable are uncorrelated, then convert the fitted model to the model of treatment effect (2.2). For instance, (2.1) can be fitted using the WLS approach because the observed measures of the treatment risk and the baseline risk are conditionally independent as they are computed based on different samples. Alternatively, the WLS approach is proposed to the linear model where the dependent variable and independent variable are the risk difference and the average of the measures of treatment risk and control risk, respectively. However, even if the correlation between the errors is avoided by choosing an appropriate model, the WLS method does not take into account the fact that the measures of treatment risk and baseline risk are observed with error since they are computed based on samples included in the meta-analysis. The inappropriate consideration of estimation error can result in biased estimators of the regression coefficients, which is discussed further in the next section.

2.2 Measurement error models

2.2.1 Measurement error

Since the measure of control risk is estimated with a summary of responses from a control group, it has measurement error. In linear regression models, failure to account for measurement error in independent variables can attenuate regression coefficients toward zero. A substantial amount of literature on measurement error in linear and nonlinear models illustrate the risk of such bias (e.g., Carroll *et al.*, 2006, Yi *et al.*, 2021). Control risk regression can lead to biased inference unless measurement error is properly accounted for, which is illustrated in van Houwelingen *et al.* (2002), Guolo *et al.* (2021) and Guolo (2021).

In control risk regression, there are two ways to model measurement errors. An approximate model directly specifies the conditional distribution of the observed measures of risk given the true ones through a bivariate normal distribution, while exact models determine the relationship between the observed risk measures and the true ones by specifying the distribution of outcomes given the true risk measures (McIntosh, 1996, Schmid *et al.*, 1998, van Houwelingen *et al.*, 2002).

2.2.2 Exact measurement error model

The choice of exact measurement error model depends on the type of data and on the risk measure. Let Y_i and X_i denote the outcomes in the treatment group and the control group in study i, respectively. In the following, we discuss exact measurement error models for continuous data, binary data and count data.

In continuous data, each study reports the mean responses $(Y_i, X_i)^\top$, the standard deviations of responses $(SD_{iT}, SD_{iC})^\top$ and the group sizes $(n_{iT}, n_{iC})^\top$. If the true risk measures η_i and ξ_i are the true mean responses, an exact measurement error model can assume a normal distribution for each mean, as follows,

$$Y_i|\eta_i \sim N\left(\eta_i, \frac{SD_{iT}^2}{n_{iT}}\right), \quad X_i|\xi_i \sim N\left(\xi_i, \frac{SD_{iC}^2}{n_{iC}}\right). \tag{2.3}$$

In binary data, each participant does or does not experience the event of interest, e.g., death. Let Y_i and X_i be the number of subjects with events. Let n_{iT} and n_{iC} be the size of the treatment group and the control group, respectively. If the risk measure is a log odds of an event, a binomial distribution can be assumed for each outcome given the true measures of risk, as follows, (Thompson *et al.*, 1997, Arends *et al.*, 2000, Schmid *et al.*, 2004)

$$Y_i|\eta_i \sim \text{Binomial}\left\{n_{iT}, \text{expit}(\eta_i)\right\}, \quad X_i|\xi_i \sim \text{Binomimal}\left\{n_{iC}, \text{expit}(\xi_i)\right\}, \tag{2.4}$$

where $\text{expit}\,(x) = \exp\,(x) \,/\, \{1 + \exp\,(x)\}$.

In count data, each study reports the number of events Y_i and X_i and the number of person-years T_{iT} and T_{iC}. If the true risk measures η_i and ξ_i are the log rate of participants with the event of interest in the treatment group and the control group, respectively, an exact measurement error model can assume a Poisson distribution for each count given the true risk measures (Arends *et al.*, 2000)

$$Y_i|\eta_i \sim \text{Poisson}\left\{\exp(\eta_i)T_{iT}\right\}, \quad X_i|\xi_i \sim \text{Poisson}\left\{\exp(\xi_i)T_{iC}\right\}. \tag{2.5}$$

2.2.3 Approximate measurement error model

Let $\hat{\eta}_i$ and $\hat{\xi}_i$ denote the observed measures of the treatment risk and of the control risk, respectively. Let $s^2_{\eta_i}$ and $s^2_{\xi_i}$ denote the within-study variances of $\hat{\eta}_i$ and $\hat{\xi}_i$, respectively, which are assume to be known and equal to their estimates. The approximate model assumes that the observed risk measures are distributed as a bivariate normal variable conditionally on the true unobserved η_i and ξ_i (McIntosh, 1996),

$$\left(\begin{array}{c} \hat{\eta}_i \\ \hat{\xi}_i \end{array} \right) \Bigg| \left(\begin{array}{c} \eta_i \\ \xi_i \end{array} \right) \sim N_2 \left\{ \left(\begin{array}{c} \eta_i \\ \xi_i \end{array} \right), \Gamma_i = \left(\begin{array}{cc} s^2_{\eta_i} & 0 \\ 0 & s^2_{\xi_i} \end{array} \right) \right\}. \tag{2.6}$$

The observed risk measures are uncorrelated since they are computed based on different groups of participants. The expression of the observed risk measures and the associated covariance matrix depends on the type of data and the choice of the risk measure. In continuous data, the observed measures of risk are the means of responses $\left(\hat{\eta}_i, \hat{\xi}_i\right)^\top = (Y_i, X_i)^\top$ and the formula of Γ_i is straightforward. In case of binary data, if the true risk measure is a log odds of an event, the observed measures of risk are (Arends *et al.*, 2000, Schmid *et al.*, 2004)

$$\hat{\eta}_i = \log\left(\frac{Y_i}{n_{iT} - Y_i}\right), \quad \hat{\xi}_i = \log\left(\frac{X_i}{n_{iC} - X_i}\right), \tag{2.7}$$

with

$$s^2_{\eta_i} = \frac{1}{Y_i} + \frac{1}{n_{iT} - Y_i}, \quad s^2_{\xi_i} = \frac{1}{X_i} + \frac{1}{n_{iC} - X_i}. \tag{2.8}$$

In case of count data, if the true risk measure is a log odds of an event, the observed measures of risk and their associated within-study variances are (Arends *et al.*, 2000)

$$\hat{\eta}_i = \log\left(\frac{Y_i}{T_{iT}}\right), \quad \hat{\xi}_i = \log\left(\frac{X_i}{T_{iC}}\right), \quad s^2_{\eta_i} = \frac{1}{Y_i}, \quad s^2_{\xi_i} = \frac{1}{X_i}.$$

In discrete data, an ad hoc correction can be applied to an observed risk measure and its within-study variance to ensure that they are well-defined. For example, 1/2 can be added to the nominator and the denominator in the formula of $\hat{\eta}_i$ and the denominators in the formula of $s^2_{\eta_i}$ when there are no participants with or without event in the treatment group of study i.

2.3 Measurement error correction

In control risk regression, there are two approaches to correct measurement errors, namely the structural approach and the functional approach. In the first way, it is assumed that the true measure of control risk is a random variable whose distribution is usually parametric and specified with unknown parameters. The functional method requires no distributional assumptions for the true measures of baseline risk, i.e., they are treated as nuisance parameters. Therefore, inference procedures are derived to address problems arising from these two approaches accordingly.

2.3.1 Structural approaches

The structural approach assumes a distribution $\mathcal{D}(\theta_\xi)$ with parameter vector θ_ξ for the true measure of control risk. Thus, the parameter of interest in this case is $\theta = (\beta_0, \beta_1, \tau^2, \theta_\xi)^\top$. Let f_{ξ_i} denote the density function of the true baseline risk. Let $f_{\eta_i|\xi_i}$ denote the conditional density function of the true treatment risk given the true control risk. Note that $\eta_i|\xi_i \sim N(\beta_0 + \beta_1\xi_i, \tau^2)$ under model (2.1). Let $f_{\hat{\eta}|\eta_i}$ and $f_{\hat{\xi}_i|\xi_i}$ denote the conditional density function of the observed risk measure given the true one in the treatment group and the control group, respectively. Under the approximate measurement error model, the likelihood function for θ can be derived by marginalizing the joint density of the observed risk measures and the true ones over the distribution of the true risk measures, as follows,

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^{n} f_{\hat{\eta}_i,\hat{\xi}_i} \\ &\propto \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\hat{\eta}_i,\hat{\xi}_i,\eta_i,\xi_i} d\eta_i d\xi_i \\ &\propto \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\hat{\eta}_i|\eta_i} f_{\hat{\xi}_i|\xi_i} f_{\eta_i|\xi_i} f_{\xi_i} d\eta_i d\xi_i, \end{aligned} \tag{2.9}$$

Let $f_{Y_i|\eta_i}$ and $f_{X_i|\xi_i}$ denote the conditional density functions of the outcomes given the true risk measures. The same technique can be applied to derive a likelihood function

if an exact measurement error is adopted

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^{n} f_{Y_i,X_i} \\ &\propto \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y_i,X_i,\eta_i,\xi_i} d\eta_i d\xi_i \\ &\propto \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y_i|\eta_i} f_{X_i|\xi_i} f_{\eta_i|\xi_i} f_{\xi_i} d\eta_i d\xi_i, \end{aligned} \tag{2.10}$$

The approximate measurement error model can result in a likelihood function which has a closed-form expression. This is however not the same for exact models in case of discrete data. Although it is more computationally convenient to perform inference with the approximate model, the distribution of the observed risk measures assumed by the approximate model is just an asymptotic distribution and hence may not be accurate when there are not sufficiently many participants in each of included groups. Moreover, when data is discrete, there is a correlation between the observed risk measures and their associated within-study variances, which may bias inference (Stijnen *et al.*, 2021, Guolo *et al.*, 2021).

For computational convenience, a normal distribution is the first choice for the baseline risk distribution (McIntosh, 1996, van Houwelingen *et al.*, 2002). Under the approximate measurement error model, the observed measures of risk marginally follow a bivariate normal distribution

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim N_2 \left\{ \begin{pmatrix} \beta_0 + \beta_1\mu_\xi \\ \mu_\xi \end{pmatrix}, \begin{pmatrix} s^2_{\eta_i} + \beta_1^2\sigma^2_\xi + \tau^2 & \beta_1\sigma^2_\xi \\ \beta_1\sigma^2_\xi & s^2_{\xi_i} + \sigma^2_\xi \end{pmatrix} \right\},$$

and the associated likelihood function is, up to constants,

$$L(\theta) \propto \prod_{i=1}^{n} \left\{\det(\Sigma + \Gamma_i)\right\}^{\frac{1}{2}} \exp\left[-\frac{1}{2} \left\{ \begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} - \boldsymbol{\mu} \right\}^{\top} (\Sigma + \Gamma_i)^{-1} \left\{ \begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} - \boldsymbol{\mu} \right\} \right],$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \beta_0 + \beta_1\mu_\xi \\ \mu_\xi \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \beta_1^2\sigma^2_\xi + \tau^2 & \beta_1\sigma^2_\xi \\ \beta_1\sigma^2_\xi & \sigma^2_\xi \end{pmatrix}.$$

The ML estimator of θ can be computed using iterative approaches or the EM algorithm (McIntosh, 1996). The EM algorithm is based on the idea that if the true measures of risk are observed, their sample means and covariance matrix are sufficient

statistics. Therefore, this algorithm estimates the sufficient statistics based on the current estimates of θ and then estimates θ based on the obtained sufficient statistics, until convergence. Alternatively, Bayesian inference can be performed with non-informative priors (McIntosh, 1996). The data augmentation algorithm can be used to compute the Bayesian estimator of θ since the posterior distribution does not have a closed-form expression. Specifically, this algorithm simulates the true risk measures based on the current values of θ and then samples this parameter vector from the updated posterior density, until convergence.

The normal specification of the control risk distribution is commonly adopted since it results in a likelihood function with closed-form expression. Furthermore, it is difficult to verify normality assumption for baseline risk because of the typically small number of studies at hand. However, the true baseline risk distribution may not be normal as a consequence of case-control scheme (Guolo, 2008). Specifically, the distribution of the true measure of control risk might be bimodal or skewed. Therefore, normality assumption for control risk might suffer from the risk of misspecification. This misspecification might not strongly affect the estimation of the regression coefficients but the estimation of the residual variance and hence might make the inference of the coefficients unreliable (Ghidey *et al.*, 2007, Guolo, 2013). As a solution, the literature has proposed several ways to increase the flexibility of the baseline risk distribution.

Reasoning that the control risk distribution could be a mixture of low- and high- risk populations, Arends *et al.* (2000) assume a mixture of two normal distributions with the same variance for the baseline risk, namely,

$$\xi_i \sim pN\left(\mu_{\xi 1}, \sigma_\xi^2\right) + (1-p)N\left(\mu_{\xi 2}, \sigma_\xi^2\right),$$

where p denotes the mixture weight to be estimated. This choice covers a wide range of distributions, including unimodal and bimodal distributions, symmetric as well as very skewed distributions. Also, under this specification of control risk distribution and exact measurement error models, Arends *et al.* (2000) perform inference according to a Bayesian approach with uninformative priors since there is no closed-form expression for the resulting likelihood functions.

A more general way to relax normality assumption for the control risk is suggested in Ghidey *et al.* (2007) which considers a finite mixture of normal distributions with the same prespecified variance

$$\xi_i \sim \sum_{j=1}^{J} \frac{\exp(a_j)}{\sum_j \exp(a_j)} N\left(\mu_{\xi j}, \sigma_\xi^2\right),$$

where $\exp(a_j)/\sum_{j=1}^{J}\exp(a_j)$ are unknown mixture weights and J denotes the tuning parameter. Means $\mu_{\xi j}$ are equally spaced on the range of possible values of the control risk. The tuning parameter J controls the trade-off between the goodness of fit and the smoothness of the control risk distribution. Large values of J make this distribution better fit the data but less smooth, while small values may yield underfitting. To optimize the smoothness and the goodness of fit, Ghidey *et al.* (2007) suggest to fix J to a large value and consider a penalized likelihood function which has a closed-form expression if being combined with the approximate measurement error model. Penalty terms avoid the large variability in the mixture weights of adjacent normal distributions and hence make the control risk distribution smooth. Given the penalty coefficient, the estimation can be easily carried out with an EM algorithm since its estimation step and maximization step are involved with two mutually independent objective functions. The penalty coefficient controls the importance of the penalty terms and hence the smoothness of the control risk distribution. This coefficient is optimized using Akaike information criterion before estimation. Through simulation, Ghidey *et al.* (2007) show that their proposed control risk distribution has slightly smaller mean squared errors of the regression coefficients compared to the normal control risk distribution when the within-study covariance matrix has small elements and the true distribution is heavily tailed.

Since the choice of the number of component distributions in the mixture is complex and still somewhat arbitrary, Lee and Thompson (2008) propose to consider a family of skew normal distributions or Student t-distributions. Let $t\left(\gamma_\xi, \omega_\xi, k\right)$ denote a Student t variable with location parameter γ_ξ, scale parameter ω_ξ and k degrees of freedom. Let f denote the associated density function. A skew Student t-distribution has the following density function

$$f_{\xi_i} = \frac{2}{\alpha_\xi + 1/\alpha_\xi}\left\{f\left(\frac{\xi_i}{\alpha_\xi}\right) I_{[0,\infty)}(\xi_i) + f\left(\alpha_\xi \xi_i\right) I_{(-\infty,0]}(\xi_i)\right\},$$

where $I_{[0,\infty)}$ denotes the indicator function of the interval $[0,\infty)$ and α_ξ denotes the parameter which controls the mass on each side of zero. Since the resulting likelihood function does not have a closed-form expression and continuous even derivatives as the proposed distribution comprises of two half distributions, the Bayesian approach can be applied. However, the proposed control risk distribution is parameterized in terms of its mode, so its mean is a complex function of the other parameters and difficult to explain (Fernandez and Steel, 1998).

An alternative way to introduce skewness to the normal control risk distribution is

proposed in Guolo (2013) based on the skew normal distribution in Azzalini (1985). This distribution is denoted with $SN(\gamma_\xi, \omega_\xi, \alpha_\xi)$, where γ_ξ is the location parameter, ω_ξ is the scale parameter, and α_ξ is the skewness parameter. The associated density function is, as follows,

$$f_{\xi_i} = \frac{2}{\omega_\xi}\phi\left(\frac{\xi_i - \gamma_\xi}{\omega_\xi}\right)\Phi\left\{\frac{\alpha_\xi(\xi_i - \gamma_\xi)}{\omega_\xi}\right\},$$

where ϕ denotes the density function and Φ denotes the cumulative distribution function of a normal distribution. As a result, the normal distribution is a special case of the skew normal distribution where the skewness parameter is equal to zero. The associated likelihood function can be derived by integrating the joint density of $\left(\hat{\eta}_i, \hat{\xi}_i, \xi_i\right)$ with respect to ξ_i (Guolo, 2013). Since there is no closed-form expression for the likelihood function, it can be approximated by numerical integration such as Gaussian-Hermite quadrature. For inference, (Guolo, 2013) suggests to use the sandwich estimator of the standard error to account for the possible misspecification of the control risk distribution. Through simulation, she shows that the proposed control risk distribution results in better maximized likelihood values and Akaike information criterion compared to the normal control risk distribution. She also shows that a skew normal distribution reduces the MSE of the estimator of the residual variance, especially when the true control risk distribution deviates much from normality assumption.

2.3.2 Functional approaches

Unlike structural measurement error correction, a functional approach for measurement error correction considers the true measures of baseline risk as nuisance parameters in place of random variables. Accordingly, this approach results in inference that is robust to the misspecification of the control risk distribution. Under the approximate measurement error model, the likelihood function for the parameter $\theta = (\beta_0, \beta_1, \tau^2, \xi_1, \cdots, \xi_n)^\top$ can be derived by integrating the joint distribution of $\left(\hat{\eta}_i, \hat{\xi}_i, \eta_i\right)^\top$ with respect to η_i

$$L(\theta) \propto \prod_{i=1}^{n} f_{\hat{\eta}_i, \hat{\xi}_i} \propto \prod_{i=1}^{n} \int_{-\infty}^{\infty} f_{\hat{\eta}_i, \hat{\xi}_i, \eta_i} d\eta_i \propto \prod_{i=1}^{n} \int_{-\infty}^{\infty} f_{\hat{\eta}_i, \hat{\xi}_i | \eta_i} f_{\eta_i} d\eta_i,$$

where $f_{\hat{\eta}_i,\hat{\xi}_i|\eta_i}$ denotes the density function in the approximate measurement error model (2.6). Since the treatment risk is normally distributed $\eta_i \sim N\left(\beta_0 + \beta_1\xi_i, \tau^2\right)$, the observed measures of risk have a bivariate normal distribution

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim N_2\left(\mu_i, \Sigma + \Gamma_i\right),$$

and the associated likelihood function can be simplified as

$$L\left(\theta\right) \propto \prod_{i=1}^{n} \left\{\det\left(\Sigma + \Gamma_i\right)\right\}^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left\{\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} - \mu_i\right\}^{\top} \left(\Sigma + \Gamma_i\right)^{-1} \left\{\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} - \mu_i\right\}\right],$$

where

$$\mu_i = \begin{pmatrix} \beta_0 + \beta_1\xi_i \\ \xi_i \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \tau^2 & 0 \\ 0 & 0 \end{pmatrix}.$$

A major disadvantage of the functional correction is that the number of parameters is of the same order as the number of studies, which makes the maximum likelihood estimators of $\left(\beta_0, \beta_1, \tau^2, \xi_1, \cdots, \xi_n\right)^{\top}$ inconsistent and the associated standard errors go wrong (Ghidey *et al.*, 2013).

Under the approximate measurement error model, the profile log-likelihood function for $\left(\beta_0, \beta_1, \tau^2\right)^{\top}$ is derived by maximizing the likelihood function with respect to the nuisance parameters ξ_i, as follows,

$$\ell_P\left(\beta_0, \beta_1, \tau^2\right) \propto -\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{\left(\hat{\eta}_i - \beta_0 - \beta_1\hat{\xi}_i\right)^2}{\tau^2 + s_{\eta_i}^2 + \beta_1^2 s_{\xi_i}^2} + \log\left(\tau^2 + s_{\eta_i}^2\right) + \log s_{\xi_i}^2\right\}.$$

The estimators of regression coefficients from the profile log-likelihood approach is consistent if the residual variance is known (Ghidey *et al.*, 2013).

Alternatively, the restricted likelihood function is obtained by integrating the likelihood function with respect to the nuisance parameters

$$\ell_{REML}\left(\beta_0, \beta_1, \tau^2\right) \propto -\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{\left(\hat{\eta}_i - \beta_0 - \beta_1\hat{\xi}_i\right)^2}{\tau^2 + s_{\eta_i}^2 + \beta_1^2 s_{\xi_i}^2} + \log\left(\tau^2 + s_{\eta_i}^2 + \beta_1^2 s_{\xi_i}^2\right)\right\}.$$

Although the REML estimator of the residual variance is consistent if regression coefficients are known, the estimators of regression coefficients are not consistent even if the

residual variance is known (Ghidey *et al.*, 2013).

Thompson *et al.* (1997) suggest a Bayesian approach with independent non-informative priors. Specifically, they choose a normal prior distribution with the fixed mean and large variance for the nuisance parameters. However, van Houwelingen and Senn (1999), Arends *et al.* (2000) shows that this approach is not good since asymptotically the resulting posterior of the parameters of interest $(\beta_0, \beta_1, \tau^2)$ is equivalent to the likelihood of an incorrectly specified model.

Walter (1997) considers model (2.1), the approximate measurement error model, and assumes no heterogeneity across studies, i.e., $\tau^2 = 0$. The maximum likelihood estimates of the regression coefficients can be simultaneously computed using iterative algorithms, for example, the Newton-Raphson algorithm. When the group sizes are similar within study and the true measure of treatment effect is small, he assumes $s^2_{\eta_i} = s^2_{\xi_i} = s^2_i$ for all i and obtains closed-form expressions for the estimators of the regression coefficients. However, Bernsen *et al.* (1999) show that the asymptotic covariance matrix of the maximum likelihood estimators of the regression coefficients in Walter (1997) is not correct. Also, Sharp and Thompson (2000) show that the proposed approach fails to account for the residual variance and the assumption for the close-form solution may not hold in practice.

Consider binary data and let a log odds of an event be the risk measure, Cook and Walter (1997) assume model (2.2) without heterogeneity across studies and a exact measurement error model similar to model (2.4). Specifically, their model is

$$\begin{aligned} \eta_i^\star &= \beta_0 + \beta_1 \xi_i, \\ Y_i | \eta_i^\star &\sim \text{Binomial}\left\{n_{iT}, \text{expit}(\eta_i^\star + \xi_i)\right\}, \\ X_i | \xi_i &\sim \text{Binomimal}\left\{n_{iC}, \text{expit}(\xi_i)\right\}. \end{aligned}$$

They obtain a likelihood function for $(\beta_0, \beta_1, \xi_1, \cdots, \xi_n)^\top$ which is a product of binomial density functions. The maximum likelihood estimates can be computed using iterative algorithms such as the Newton-Raphson method. Although an advantage of the proposed method over the one in Walter (1997) is the ability to properly account for within-study covariances, it still neither resolves the aforementioned inconsistency issue nor accounts for the residual variance.

Under model (2.1) and the approximate measurement error model (2.6), Ghidey *et al.* (2013) propose corrected score equations and conditional score equations which are based on Carroll *et al.* (2006) to estimate parameter vector $(\beta_0, \beta_1, \tau^2)^\top$. To derive corrected score equations, they apply unbiased estimating equations in linear regression

with additive measurement error in covariates (Carroll *et al.*, 2006) to account for the presence of measurement error in the control risk and then modify these equations to consider the variation across within-study variances of the treatment risk. As an alternative to corrected score equations, they first construct a minimum variance unbiased estimator of ξ_i based on its sufficient statistics. Then they develope conditional score equations based on the first two conditional moments of $\hat{\eta}_i$ given the obtained estimator of ξ_i. The corrected score equations and conditional score equations are solved by iterative algorithms such as the Newton-Raphson method. They both yield consistent estimators of the parameters of interest since they are unbiased estimating equations. A sandwich estimator can be used for the standard error to account for possible misspecification. Through simulation, while the two proposed methods perform similarly when the within-study variance of the control risk is small, the conditional estimating equations are more efficient when the within-variance is large and the number of studies is small. However, the conditional estimating equations have multiple roots not all of which are consistent (Stefanski and Carroll, 1987, Tsiatis and Davidian, 2001).

Under models (2.1) and (2.6), Guolo (2014) suggests adapting a simulation extrapolation (SIMEX) approach which is a simulation-based approach to estimate the parameters of interest and reduce bias due to measurement errors (Carroll *et al.*, 2006). Although this method is first derived in Cook and Stefanski (1994), Stefanski and Cook (1995) for additive measurement errors, it can be well applied on any types of measurement error which can be simulated via Monte Carlo approaches. The idea behind the method is that the effect of measurement errors on an estimator can be determined experimentally via simulation. SIMEX is carried out into two steps. In the first step, resampling-like strategies are used to create additional datasets with increasing measurement error. Each of the obtained datasets provides an estimate of the parameter vector by, for example, using the naive approach. In the second step, the relationship between the obtained estimates and the additional datasets is evaluated and used to extrapolate the corrected (SIMEX) estimate to the case of no error. Since the idea behind SIMEX is simple and its application is straightforward, this approach is widespread in application. Nevertheless, the computational effort required by SIMEX is expected to increase in case of multiple covariates. See the discussion in Guolo (2014) and Guolo *et al.* (2021).

Chapter 3

Control risk regression with additional covariates

3.1 Error-free covariates

Suppose that all the included studies provide information about characteristics relating to the studies' design or patients, useful to explain heterogeneity. Boissel *et al.* (2008) discuss the possibility to add characteristics which are not accounted in the baseline risk to the control risk regression model. Let encode a characteristic of interest by a covariate ζ_i.
If ζ_i is measured or observed without error and there is no risk of aggregation bias from this covariate, model (2.1) can be extended to include ζ_i, as follows,

$$\eta_i = \beta_0 + \beta_1 \xi_i + \beta_2 \zeta_i + \varepsilon_i, \quad \xi_i \sim N\left(\mu_\xi, \sigma_\xi^2\right), \quad \varepsilon_i \sim N\left(0, \tau^2\right). \tag{3.1}$$

The error-free covariate can be continuous, e.g., the year when a study was conducted, or discrete, e.g., the place where that study was conducted. The inclusion of ζ_i does not lead to any modification of measurement error model (2.6). Therefore, the marginal distribution of the observed measures of risk is a bivariate normal distribution whose mean vector is slightly different from the mean vector in the classical control risk regression model, i.e.,

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \end{pmatrix} \sim N_2 \left\{ \begin{pmatrix} \beta_0 + \beta_1 \mu_\xi + \beta_2 \zeta_i \\ \mu_\xi \end{pmatrix}, \begin{pmatrix} s_{\eta_i}^2 + \beta_1^2 \sigma_\xi^2 + \tau^2 & \beta_1 \sigma_\xi^2 \\ \beta_1 \sigma_\xi^2 & s_{\xi_i}^2 + \sigma_\xi^2 \end{pmatrix} \right\}.$$

In this case, the parameter vector if interest θ is $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \tau^2, \sigma_\xi^2\right)^\top$. In the rest of the chapter, the focus will be on error-affected covariates.

3.2 Error-affected covariates

The additional covariate ζ_i is affected by measurement error when it represents aggregated information from all the subjects included in each study, as, for example, the summary information about the age of individuals. Consider ζ_i following a classical measurement error model, or an additive measurement error model, according to the terminology in Carroll *et al.* (2006). The erroneous measures of ξ_i and ζ_i require the specification of the distributions of the true measures, which can be assumed normal and independent for computational convenience and in analogy to what is commonly done in the classical control risk regression model. The model of η_i, ξ_i and ζ_i is thus

$$\eta_i = \beta_0 + \beta_1 \xi_i + \beta_2 \zeta_i + \varepsilon_i, \tag{3.2}$$

where

$$\xi_i \sim N\left(\mu_\xi, \sigma_\xi^2\right), \quad \zeta_i \sim N\left(\mu_\zeta, \sigma_\zeta^2\right), \quad \varepsilon_i \sim N\left(0, \tau^2\right).$$

In practice, the assumption of independence between ξ_i and ζ_i in model (3.2) may be questionable. For example, consider a covariate reporting the mean age of patients affected by a pathology that is known to mainly affect the elderly or a covariate reporting the log odds of male among patients affected by a pathology that is known to mainly affect men. Let Z_i denote the outcome and $\hat{\zeta}_i$ denote the observed value of covariate ζ_i. Model (3.2) can be accompanied by an extension of the approximate measurement error model (2.6). Thus,

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \\ \hat{\zeta}_i \end{pmatrix} \Bigg| \begin{pmatrix} \eta_i \\ \xi_i \\ \zeta_i \end{pmatrix} \sim N_3 \left\{ \begin{pmatrix} \eta_i \\ \xi_i \\ \zeta_i \end{pmatrix}, \Gamma_i = \begin{pmatrix} s_{\eta_i}^2 & 0 & s_{\eta_i,\zeta_i} \\ 0 & s_{\xi_i}^2 & s_{\xi_i,\zeta_i} \\ s_{\eta_i,\zeta_i} & s_{\xi_i,\zeta_i} & s_{\zeta_i}^2 \end{pmatrix} \right\}, \tag{3.3}$$

where $s_{\zeta_i}^2$ denotes the within-study variance of $\hat{\zeta}_i$, s_{η_i,ζ_i} denotes the within-study covariance between $\hat{\eta}_i$ and $\hat{\zeta}_i$, and s_{ξ_i,ζ_i} denotes the within-study covariance between $\hat{\xi}_i$ and $\hat{\zeta}_i$.

Under the assumption of independence between ξ_i and ζ_i, the marginal distribution of the observed measures of risk and the covariate $\left(\hat{\eta}_i, \hat{\xi}_i, \hat{\zeta}_i\right)^\top$ in study i is

$$\begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \\ \hat{\zeta}_i \end{pmatrix} \sim N_3\left(\boldsymbol{\mu}, \Sigma + \Gamma_i\right), \tag{3.4}$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \beta_0 + \beta_1\mu_\xi + \beta_2\mu_\zeta \\ \mu_\xi \\ \mu_\zeta \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \beta_1^2\sigma_\xi^2 + \beta_2^2\sigma_\zeta^2 + \tau^2 & \beta_1\sigma_\xi^2 & \beta_2\sigma_\zeta^2 \\ \beta_1\sigma_\xi^2 & \sigma_\xi^2 & 0 \\ \beta_2\sigma_\zeta^2 & 0 & \sigma_\zeta^2 \end{pmatrix}.$$

In this case, the whole parameter vector becomes $\theta = \left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and the associated likelihood function has a closed-form expression,

$$\begin{aligned} L(\theta) &\propto \prod_{i=1}^{n} f_{\hat{\eta}_i,\hat{\xi}_i,\hat{\zeta}_i} \\ &\propto \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\hat{\eta}_i,\hat{\xi}_i,\hat{\zeta}_i|\eta_i,\xi_i,\zeta_i} \times f_{\eta_i|\xi_i,\zeta_i} \times f_{\xi_i} \times f_{\zeta_i} d\eta_i d\xi_i d\zeta_i \\ &\propto \prod_{i=1}^{n} \left\{\det\left(\Sigma + \Gamma_i\right)\right\}^{\frac{1}{2}} \phi_3 \left[\left(\Sigma + \Gamma_i\right)^{-\frac{1}{2}} \left\{ \begin{pmatrix} \hat{\eta}_i \\ \hat{\xi}_i \\ \hat{\zeta}_i \end{pmatrix} - \boldsymbol{\mu} \right\} \right], \end{aligned} \tag{3.5}$$

where ϕ_3 denotes the density function of a three-dimensional standard normal variable. As in the control risk regression model without additional covariates, the within-study variance of $\hat{\zeta}_i$, $s_{\zeta_i}^2$, can be assumed known and equal to its estimator $\widehat{var}\left(\hat{\zeta}_i|\zeta_i\right)$, provided that the sample size in every study is large enough. However, the within-study variance might be unknown if the information about the additional covariate is only given in terms of the mean of values at the individual level. In this case, this within-study variance can be substituted with appropriate values based on prior or experts' knowledge.

Within-study covariances s_{η_i,ζ_i} and s_{ξ_i,ζ_i} can be computed if the amount of information about the additional covariate is sufficient, e.g., if individual participant data (IPD) are available. Therefore, in this case, it is possible to compute the associated likelihood function using the sample estimators of within-study covariances. While the case of available IPD is not common, subgroup summary information is more likely to be available and it still allows to estimate within-study covariances. Likelihood functions based on subgroup summary information will be proposed in the following subsections. Without loss of generality, in this chapter, let us assume Y_i and X_i are the number of subjects with event and the number of subjects without event, and the risk measure is a log odds of an event.

Treatment group				Control group			
Event		Non-event		Event		Non-event	
Male	Female	Male	Female	Male	Female	Male	Female
Z_{iT1M}	Z_{iT1F}	Z_{iT0M}	Z_{iT0F}	Z_{iC1M}	Z_{iC1F}	Z_{iC0M}	Z_{iC0F}

TABLE 3.1: Subgroup summary for covariate ζ_i expressed as a log odds.

3.2.1 Covariate ζ_i expressed as a log odds

Suppose that every study provides the number of subjects in subgroups specified with the treatment/control group, an event of interest (e.g., death) and a category of interest (e.g., male individuals). Therefore, subgroup summary information is given in terms of the sizes of subgroups and is shown in Table 3.1, where the subscripts 'M' and 'F' indicate male and female, respectively, the subscripts 'T' and 'C' indicate the treatment group and the control group, respectively, and the subscripts 1 and 0 indicate 'with event' and 'without event', respectively. For example, Z_{iT1M} denotes the number of treated male patients with event. Let Z_i denote the number of male patients in study i, so $Z_i = Z_{iT1M} + Z_{iT0M} + Z_{iC1M} + Z_{iC0M}$. Assume that the male indicator at the individual level is identically distributed as a Bernoulli variable with probability $\text{expit}(\zeta_i)$, so Z_i has a binomial distribution

$$Z_i|\zeta_i \sim \text{Binomial}\left\{n_i, \text{expit}(\zeta_i)\right\}, \tag{3.6}$$

where n_i denotes the sample size in study i, i.e., $n_i = n_{iT} + n_{iC}$. An estimator of ζ_i an its within-study variance respectively are

$$\hat{\zeta}_i = \log\left(\frac{Z_i}{n_i - Z_i}\right), \quad s^2_{\zeta_i} = \frac{1}{Z_i} + \frac{1}{n_i - Z_i}.$$

In this case, it is possible to compute within-study covariances using the first-order Taylor's expansion around the means (Bagos, 2012). Thus,

$$\begin{aligned}
s_{\eta_i,\zeta_i} &= cov\left(\hat{\eta}_i, \hat{\zeta}_i\right) \\
&= cov\left(\log\frac{Y_i}{n_{iT}-Y_i}, \log\frac{Z_i}{n_i-Z_i}\right) \\
&= cov\left(\log Y_i, \log Z_i\right) - cov\left\{\log(n_{iT}-Y_i), \log Z_i\right\} \\
&\quad -cov\left\{\log Y_i, \log(n_i-Z_i)\right\} + cov\left\{\log(n_{iT}-Y_i), \log(n_i-Z_i)\right\} \\
&\approx \frac{1}{E\left(Y_i\right)E\left(Z_i\right)}cov\left(Y_i, Z_i\right) - \frac{1}{E(n_{iT}-Y_i)E\left(Z_i\right)}cov\left(n_{iT}-Y_i, Z_i\right) \\
&\quad -\frac{1}{E\left(Y_i\right)E(n_i-Z_i)}cov\left(Y_i, n_i-Z_i\right) \\
&\quad +\frac{1}{E(n_{iT}-Y_i)E(n_i-Z_i)}cov\left(n_{iT}-Y_i, n_i-Z_i\right),
\end{aligned}$$

where *cov* denotes 'within-study covariance' or 'covariance given $(\eta_i, \xi_i, \zeta_i)^\top$'. The expected values of Y_i and Z_i can be estimated with their observed values. Let U_{ik} and V_{ik}, $k = 1, \cdots, n$ denote the indicator of event and the indicator of male of subject k in the treatment group of study i, respectively. The within-study covariance between Y_i and Z_i can be estimated as

$$\begin{aligned}
cov\left(Y_i, Z_i\right) &= cov\left(Y_i, Z_{iT1M} + Z_{iT0M} + Z_{iC1M} + Z_{iC0M}\right) \\
&= cov\left(Y_i, Z_{iT1M} + Z_{iT0M}\right) \\
&= \sum_{j=1}^{n_{iT}}\sum_{k=1}^{n_{iT}} cov\left(U_{ij}, V_{ik}\right) \\
&= \sum_{k=1}^{n_{iT}} cov\left(U_{ik}, V_{ik}\right) \\
&= n_{iT}cov\left(U_{i1}, V_{i1}\right) \\
&= n_{iT}\left\{E(U_{i1}V_{i1}) - E(U_{i1})E(V_{i1})\right\} \\
&= n_{iT}\{Pr\left(\text{a treated patient is male and with event}\right) \\
&\quad - Pr(\text{a treated patient is with event})Pr(\text{a treated patient is male})\} \\
&\approx n_{iT}\left\{\frac{Z_{iT1M}}{n_{iT}} - \frac{Y_i}{n_{iT}} \times \frac{\left(Z_{iT1M} + Z_{iT0M}\right)}{n_{iT}}\right\} \\
&= Z_{iT1M} - \frac{Y_i\left(Z_{iT1M} + Z_{iT0M}\right)}{n_{iT}}.
\end{aligned}$$

Treatment group		Control group	
Event	Non-event	Event	Non-event
Z_{iT1}	Z_{iT0}	Z_{iC1}	Z_{iC0}

TABLE 3.2: Subgroup summary for covariate ζ_i expressed as a mean.

Other within-study covariances in the expansion of s_{η_i,ζ_i} can be estimated similarly,

$$\begin{aligned} cov\left(n_{iT}-Y_i, Z_i\right) &\approx Z_{iT0M} - \frac{\left(n_{iT}-Y_i\right)\left(Z_{iT1M}+Z_{iT0M}\right)}{n_{iT}}, \\ cov\left(Y_i, n_i - Z_i\right) &\approx Z_{iT1F} - \frac{Y_i\left(Z_{iT1F}+Z_{iT0F}\right)}{n_{iT}}, \\ cov\left(n_{iT}-Y_i, n_i - Z_i\right) &\approx Z_{iT0F} - \frac{\left(n_{iT}-Y_i\right)\left(Z_{iT1F}+Z_{iT0F}\right)}{n_{iT}}. \end{aligned}$$

Hence, the within-study covariance between $\hat{\eta}_i$ and $\hat{\zeta}_i$ is

$$s_{\eta_i,\zeta_i} \approx \frac{1}{Z_i}\left(\frac{Z_{iT1M}}{Y_i} - \frac{Z_{iT0M}}{n_{iT}-Y_i}\right) - \frac{1}{n_i - Z_i}\left(\frac{Z_{iT1F}}{Y_i} - \frac{Z_{iT0F}}{n_{iT}-Y_i}\right),$$

and the within-study covariance between $\hat{\xi}_i$ and $\hat{\zeta}_i$ is

$$s_{\xi_i,\zeta_i} \approx \frac{1}{Z_i}\left(\frac{Z_{iC1M}}{X_i} - \frac{Z_{iC0M}}{n_{iC}-X_i}\right) - \frac{1}{n_i - Z_i}\left(\frac{Z_{iC1F}}{X_i} - \frac{Z_{iC0F}}{n_{iC}-X_i}\right).$$

3.2.2 Covariate ζ_i expressed as the mean response of a characteristic

Suppose that every study provides the mean response of a characteristic (e.g., the mean age of patients) in subgroups specified with the treatment/control group and an event of interest (e.g., death). Therefore, subgroup summary information can be given in the form of Table 3.2 where the subscripts 'T' and 'C' indicate the treatment group and the control group, respectively, and the subscripts 1 and 0 indicate 'with event' or 'without event'. For example, Z_{iT1} denotes the mean age of treated patients with event. Let Z_i denote the mean age of patients in study i, so

$$Z_i = \frac{Y_i Z_{iT1} + \left(n_{iT}-Y_i\right) Z_{iT0} + X_i Z_{iC1} + \left(n_{iC}-X_i\right) Z_{iC0}}{n_i}.$$

Assume that the ages of individuals in study i are identically distributed as normal variables with mean ζ_i and variance SD_i^2, so Z_i has a normal distribution

$$Z_i|\zeta_i \sim N\left(\zeta_i, \frac{SD_i^2}{n_i}\right). \tag{3.7}$$

An estimator of ζ_i and its within-study variance respectively are

$$\hat{\zeta}_i = Z_i, \quad s^2_{\zeta_i} = \frac{SD_i^2}{n_i}.$$

Covariances between the observed measures of risk and covariate can be approximated using the first-order Taylor's expansion around the means (Bagos, 2012). Let Z_{iT} and Z_{iC} denote the mean age in the treatment group and in the control group, respectively. Thus,

$$\begin{aligned}
s_{\eta_i,\zeta_i} &= cov\left(\hat{\eta}_i, \hat{\zeta}_i\right) \\
&= cov\left(\log\frac{Y_i}{n_{iT}-Y_i}, Z_i\right) \\
&= cov\left(\log\frac{Y_i}{n_{iT}-Y_i}, \frac{n_{iT}Z_{iT}+n_{iC}Z_{iC}}{n_i}\right) \\
&= cov\left\{\log Y_i - \log(n_{iT}-Y_i), \frac{n_{iT}}{n_i}Z_{iT}\right\} \\
&= \frac{n_{iT}}{n_i}cov\left(\log Y_i, Z_{iT}\right) - \frac{n_{iT}}{n_i}cov\left\{\log(n_{iT}-Y_i), Z_{iT}\right\} \\
&\approx \frac{n_{iT}}{n_i}\frac{1}{E\left(Y_i\right)}cov\left(Y_i, Z_{iT}\right) - \frac{n_{iT}}{n_i}\frac{1}{E\left(n_{iT}-Y_i\right)}cov\left(n_{iT}-Y_i, Z_{iT}\right),
\end{aligned}$$

The expected value of Y_i can be estimated with its observed value. Let U_{ik} denote the indicator of event and V_{ik} denote the age of subject k in the treatment group of study i. The covariance between Y_i and Z_{iT} can be estimated as

$$\begin{aligned}
cov\left(Y_i, Z_{iT}\right) &= \frac{1}{n_{iT}}cov\left(\sum_{j=1}^{n_{iT}}U_{ij}, \sum_{k=1}^{n_{iT}}V_{ik}\right) \\
&= \frac{1}{n_{iT}}\sum_{j=1}^{n_{iT}}\sum_{k=1}^{n_{iT}}cov\left(U_{ij}, V_{ik}\right) \\
&= \frac{1}{n_{iT}}\sum_{k=1}^{n_{iT}}cov\left(U_{ik}, V_{ik}\right) \\
&= \frac{1}{n_{iT}}n_{iT}cov\left(U_{i1}, V_{i1}\right) \\
&= E\left(U_{i1}V_{i1}\right) - E\left(U_{i1}\right)E\left(V_{i1}\right).
\end{aligned}$$

Since $U_{i1}V_{i1}$ is equal to V_{i1} if treated patient 1 is with event and zero otherwise, its expected value can be estimated by $Y_i Z_{iT1}/n_{iT}$, so that

$$cov\left(Y_i, Z_{iT}\right) \approx \frac{Y_i}{n_{iT}}\left(Z_{iT1} - Z_{iT}\right).$$

The within-study covariance between $n_{iT} - Y_i$ and Z_{iT} can be estimated similarly,

$$cov\left(n_{iT} - Y_i, Z_{iT}\right) \approx \frac{n_{iT} - Y_i}{n_{iT}}\left(Z_{iT0} - Z_{iT}\right).$$

Therefore, the within-study covariance between $\hat{\eta}_i$ and $\hat{\zeta}_i$ is

$$s_{\eta_i,\zeta_i} \approx \frac{1}{n_i}\left(Z_{iT1} - Z_{iT0}\right),$$

and the within-study covariance between $\hat{\xi}_i$ and $\hat{\zeta}_i$ is

$$s_{\xi_i,\zeta_i} \approx \frac{1}{n_i}\left(Z_{iC1} - Z_{iC0}\right).$$

3.2.3 Pseudo-likelihood approach

As a typical problem of missing values in multivariate meta-analysis, within-study covariances are expected not to be available in studies included in the meta-analysis. As a result, inference can not be performed using the likelihood function (3.5). It is hence desirable to base inference on a function which is derived from the reduction of the complexity of the marginal distribution (3.4) and which maintains good likelihood properties of the estimators of the parameters of interest. Such a function is referred to as the pseudo-likelihood function (Besag, 1975). Our proposal in case of unavailable covariances is a pseudo-likelihood function which sets within-study covariances to zero

$$pL\left(\theta\right) \propto \prod_{i=1}^{n}\left\{\det\left(\Sigma + \tilde{\Gamma}_i\right)\right\}^{\frac{1}{2}} \phi_3\left[\left(\Sigma + \tilde{\Gamma}_i\right)^{-\frac{1}{2}}\left\{\begin{pmatrix}\hat{\eta}_i\\ \hat{\xi}_i\\ \hat{\zeta}_i\end{pmatrix} - \boldsymbol{\mu}\right\}\right],$$

where $\tilde{\Gamma}_i = \mathrm{diag}\left\{s^2_{\eta_i}, s^2_{\xi_i}, s^2_{\zeta_i}\right\}$. In other words, this function is derived under a working conditional independence assumption of $\hat{\eta}_i, \hat{\xi}_i$ and $\hat{\zeta}_i$. The pseudo-likelihood estimator can be computed using iterative algorithms such as the Newton-Raphson algorithm. Although the pseudo-likelihood approach reduces the complexity of the marginal distribution, there might be the risk of misspecification from the reduced model assumed by this approach, i.e., measures of risk might be correlated with the additional covariate

within study. Therefore, the sandwich matrix can be used when performing inference to account for possible misspecification (Kauermann and Carroll, 2001, Ghidey *et al.*, 2013, Guolo, 2012), which is defined as

$$G\left(\theta\right) = \frac{1}{n} J^{-1}(\theta) I(\theta) J^{-1}(\theta),$$

where

$$J\left(\theta\right) = \frac{1}{n} \frac{\partial^2 \log pL(\theta)}{\partial\theta\partial\theta^\top}, \quad I(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log pL_i(\theta)}{\partial\theta} \left(\frac{\partial \log pL_i(\theta)}{\partial\theta}\right)^\top.$$

3.2.4 Exact measurement error models for additional covariates

As for the classical control risk regression model without additional covariates, also in case of study-specific covariates, the control risk regression model can be specified using the exact measurement error description in place of the approximate version. For example, when the additional covariate is a log odds and subgroup summary information is given in terms of Table 3.1, the subgroup outcomes in Table 3.1 can be modeled using a multinomial distribution

$$\begin{aligned} \left(Z_{iT1M}, Z_{iT0M}, Z_{iT1F}, Z_{iT0F}\right)^\top \Big| \left(\eta_i, \zeta_i\right)^\top &\sim \text{Multinomial}_4\left(p_{iT1M}, p_{iT0M}, p_{iT1F}, p_{iT0F}\right), \\ \left(Z_{iC1M}, Z_{iC0M}, Z_{iC1F}, Z_{iC0F}\right)^\top \Big| \left(\xi_i, \zeta_i\right)^\top &\sim \text{Multinomial}_4\left(p_{iC1M}, p_{iC0M}, p_{iC1F}, p_{iC0F}\right), \end{aligned} \tag{3.8}$$

where vectors of probabilities $\left(p_{iT1M}, p_{iT0M}, p_{iT1F}, p_{iT0F}\right)^\top$ and $\left(p_{iC1M}, p_{iC0M}, p_{iC1F}, p_{iC0F}\right)^\top$ are functions of $\left(\eta_i, \zeta_i\right)^\top$ and $\left(\xi_i, \zeta_i\right)^\top$, respectively, as follows,

$$\begin{aligned} p_{iT1M} + p_{iT1F} &= \text{expit}(\eta_i) \\ p_{iT1M} + p_{iT0M} &= \text{expit}(\zeta_i) \\ p_{iT1M} + p_{iT0M} + p_{iT1F} + p_{iT0F} &= 1 \\ p_{iC1M} + p_{iC1F} &= \text{expit}(\xi_i) \\ p_{iC1M} + p_{iC0M} &= \text{expit}(\zeta_i) \\ p_{iC1M} + p_{iC0M} + p_{iC1F} + p_{iC0F} &= 1. \end{aligned}$$

In this case, the likelihood function for θ has a more complex expression if compared to the case of the approximate measurement error model, namely,

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{Z_{iT1M}, Z_{iT0M}, Z_{iT1F}, Z_{iT0F} | \eta_i, \zeta_i} \\ &\quad \times f_{Z_{iC1M}, Z_{iC0M}, Z_{iC1F}, Z_{iC0F} | \xi_i, \zeta_i} \times f_{\eta_i | \xi_i, \zeta_i} \times f_{\xi_i} \times f_{\zeta_i} d\eta_i d\xi_i d\zeta_i \\ &\propto \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{iT1M}^{Z_{iT1M}} \times p_{iT0M}^{Z_{iT0M}} \times p_{iT1F}^{Z_{iT1F}} \times p_{iT0F}^{Z_{iT0F}} \\ &\quad \times p_{iC1M}^{Z_{iC1M}} \times p_{iC0M}^{Z_{iC0M}} \times p_{iC1F}^{Z_{iC1F}} \times p_{iC0F}^{Z_{iC0F}} \times \frac{1}{\sqrt{\tau^2}} \times \frac{1}{\sqrt{\sigma_\xi^2}} \times \frac{1}{\sqrt{\sigma_\zeta^2}} \\ &\quad \times \phi\left(\frac{\eta_i - \beta_0 - \beta_1 \xi_i - \beta_2 \zeta_i}{\tau}\right) \phi\left(\frac{\xi_i - \mu_\xi}{\sigma_\xi}\right) \phi\left(\frac{\zeta_i - \mu_\zeta}{\sigma_\zeta}\right) d\eta_i d\xi_i d\zeta_i. \end{aligned}$$

Such a likelihood function does not have a closed-form expression and is very computational expensive to maximize, since every integrand in this function is a product of many functions involving $(\eta_i, \xi_i, \zeta_i)^\top$.

When the additional covariate is the mean response of a characteristic and subgroup summary information is given in terms of Table 3.2, the subgroup outcomes given X_i and Y_i can be modeled using bivariate normal distributions, i.e.,

$$\begin{aligned} \begin{pmatrix} Z_{iT1} \\ Z_{iT0} \end{pmatrix} \Bigg| \begin{pmatrix} Y_i \\ \zeta_i \end{pmatrix} &\sim N_2 \left[\begin{pmatrix} \zeta_i \\ \zeta_i \end{pmatrix}, \begin{Bmatrix} \frac{SD_i^2}{Y_i} & \frac{\rho_{iT} SD_i^2}{\sqrt{Y_i(n_{iT} - Y_i)}} \\ \frac{\rho_{iT} SD_i^2}{\sqrt{Y_i(n_{iT} - Y_i)}} & \frac{SD_i^2}{n_{iT} - Y_i} \end{Bmatrix} \right], \\ \begin{pmatrix} Z_{iC1} \\ Z_{iC0} \end{pmatrix} \Bigg| \begin{pmatrix} X_i \\ \zeta_i \end{pmatrix} &\sim N_2 \left[\begin{pmatrix} \zeta_i \\ \zeta_i \end{pmatrix}, \begin{Bmatrix} \frac{SD_i^2}{X_i} & \frac{\rho_{iC} SD_i^2}{\sqrt{X_i(n_{iC} - X_i)}} \\ \frac{\rho_{iC} SD_i^2}{\sqrt{X_i(n_{iC} - X_i)}} & \frac{SD_i^2}{n_{iC} - X_i} \end{Bmatrix} \right], \end{aligned} \tag{3.9}$$

where ρ_{iT} and ρ_{iC} denote within-study correlations which are functions of Y_i and X_i, respectively. Again, the associated likelihood function has a complex form,

$$\begin{aligned}
L(\theta) &= \prod_{i=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Z_{iT1},Z_{iT0}|Y_i,\zeta_i}\times f_{Y_i|\eta_i} \\
&\quad \times f_{Z_{iC1},Z_{iC0}|X_i,\zeta_i}\times f_{X_i|\xi_i}\times f_{\eta_i|\xi_i,\zeta_i}\times f_{\xi_i}\times f_{\zeta_i} d\eta_i d\xi_i d\zeta_i \\
&\propto \prod_{i=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_2\left[\begin{Bmatrix} \frac{SD_i^2}{Y_i} & \frac{\rho_{iT}SD_i^2}{\sqrt{Y_i(n_{iT}-Y_i)}} \\ \frac{\rho_{iT}SD_i^2}{\sqrt{Y_i(n_{iT}-Y_i)}} & \frac{SD_i^2}{n_{iT}-Y_i}\end{Bmatrix}^{-\frac{1}{2}}\left\{\begin{pmatrix} Z_{iT1} \\ Z_{iT0}\end{pmatrix}-\begin{pmatrix}\zeta_i \\ \zeta_i\end{pmatrix}\right\}\right] \\
&\quad \times\phi_2\left[\begin{Bmatrix} \frac{SD_i^2}{X_i} & \frac{\rho_{iC}SD_i^2}{\sqrt{X_i(n_{iC}-X_i)}} \\ \frac{\rho_{iC}SD_i^2}{\sqrt{X_i(n_{iC}-X_i)}} & \frac{SD_i^2}{n_{iC}-X_i}\end{Bmatrix}^{-\frac{1}{2}}\left\{\begin{pmatrix} Z_{iC1} \\ Z_{iC0}\end{pmatrix}-\begin{pmatrix}\zeta_i \\ \zeta_i\end{pmatrix}\right\}\right] \\
&\quad \times\left\{\text{expit}\left(\eta_i\right)\right\}^{Y_i}\left\{1-\text{expit}\left(\eta_i\right)\right\}^{n_{iT}-Y_i}\left\{\text{expit}\left(\xi_i\right)\right\}^{X_i}\left\{1-\text{expit}\left(\xi_i\right)\right\}^{n_{iC}-X_i} \\
&\quad \times\frac{1}{\sqrt{\tau^2}}\phi\left(\frac{\eta_i-\beta_0-\beta_1\xi_i-\beta_2\zeta_i}{\tau}\right)\frac{1}{\sqrt{\sigma_\xi^2}}\phi\left(\frac{\xi_i-\mu_\xi}{\sigma_\xi}\right)\frac{1}{\sqrt{\sigma_\zeta^2}}\phi\left(\frac{\zeta_i-\mu_\zeta}{\sigma_\zeta}\right)d\eta_i d\xi_i d\zeta_i,
\end{aligned}$$

where ϕ and ϕ_2 denote the density functions of the univariate standard normal distribution and the bivariate standard normal distribution, respectively.
When subgroup summary information is unavailable, the likelihood function for θ has a general expression

$$L(\theta) \propto \prod_{i=1}^{n}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y_i,X_i,Z_i|\eta_i,\xi_i,\zeta_i} f_{\eta_i|\xi_i,\zeta_i} f_{\xi_i} f_{\zeta_i} d\eta_i d\xi_i d\zeta_i,$$

where dependence of the functions in the integrand on the parameters of interest is suppressed for convenience of notation. It is much more complex to specify a conditional joint distribution of Y_i, X_i and Z_i, $f_{Y_i,X_i,Z_i|\eta_i,\xi_i,\zeta_i}$, than to specify a conditional distribution for each of these variables, namely, $f_{Y_i|\eta_i}$, $f_{X_i|\xi_i}$ and $f_{Z_i|\zeta_i}$. Therefore, in analogy to the case of the approximate measurement error model, it is possible to use a pseudo-likelihood function which assumes conditional independence between Y_i, X_i and Z_i within each study. If ζ_i is expressed as a log odds, a pseudo-likelihood function can

be derived using model (3.6),

$$\begin{aligned} pL(\theta) \;\propto\; & \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{Y_i|\eta_i} f_{X_i|\xi_i} f_{Z_i|\zeta_i} f_{\eta_i|\xi_i,\zeta_i} f_{\xi_i} f_{\zeta_i} d\eta_i d\xi_i d\zeta_i \\ \propto\; & \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{\text{expit}\,(\eta_i)\}^{Y_i} \{1-\text{expit}\,(\eta_i)\}^{n_{iT}-Y_i} \\ & \times \{\text{expit}\,(\xi_i)\}^{X_i} \{1-\text{expit}\,(\xi_i)\}^{n_{iC}-X_i} \{\text{expit}\,(\zeta_i)\}^{Z_i} \{1-\text{expit}\,(\zeta_i)\}^{n_i-Z_i} \\ & \times \frac{1}{\sqrt{\tau^2}}\phi\left(\frac{\eta_i-\beta_0-\beta_1\xi_i-\beta_2\zeta_i}{\tau}\right) \frac{1}{\sqrt{\sigma_\xi^2}}\phi\left(\frac{\xi_i-\mu_\xi}{\sigma_\xi}\right) \frac{1}{\sqrt{\sigma_\zeta^2}}\phi\left(\frac{\zeta_i-\mu_\zeta}{\sigma_\zeta}\right) d\eta_i d\xi_i d\zeta_i. \end{aligned}$$

In case of covariate expressed as the mean response of a characteristic, a pseudo-likelihood function can be derived using model (3.7),

$$\begin{aligned} pL(\theta) \;\propto\; & \prod_{i=1}^{n} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{\text{expit}\,(\eta_i)\}^{Y_i} \{1-\text{expit}\,(\eta_i)\}^{n_{iT}-Y_i} \\ & \times \{\text{expit}\,(\xi_i)\}^{X_i} \{1-\text{expit}\,(\xi_i)\}^{n_{iC}-X_i} \\ & \times \phi\left(\frac{Z_i-\zeta_i}{SD_i/\sqrt{n_i}}\right) \times \frac{1}{\sqrt{\tau^2}}\phi\left(\frac{\eta_i-\beta_0-\beta_1\xi_i-\beta_2\zeta_i}{\tau}\right) \\ & \frac{1}{\sqrt{\sigma_\xi^2}}\phi\left(\frac{\xi_i-\mu_\xi}{\sigma_\xi}\right) \frac{1}{\sqrt{\sigma_\zeta^2}}\phi\left(\frac{\zeta_i-\mu_\zeta}{\sigma_\zeta}\right) d\eta_i d\xi_i d\zeta_i. \end{aligned}$$

Since there is no closed-form expression for a pseudo-likelihood function when an exact measurement error model is assumed, the pseudo-likelihood estimator can be computed using numerical integration, as, for example, the Gaussian-Hermite quadrature. When performing inference, the sandwich standard error can be used to make the approach robust to model misspecification.

3.3 Simulation study

3.3.1 Set-up

In this section, a simulation study is carried out to evaluate the performance of the likelihood approach and the pseudo-likelihood approach which were proposed in the previous sections. These methods are used to fit the control risk regression model with an additional covariate, ζ_i to data which are generated under each of the following scenarios

Scenario 1 Model (3.1) with $\zeta_i \sim N(0,1)$ fixed;

Scenario 2 Model (3.1) with $\zeta_i \sim \text{Bernoulli}\,(0.5)$ fixed;

Scenario 3 Model (3.2) and model (3.8);

Scenario 4 Model (3.2) and model (3.9);

where $(\beta_0, \beta_1, \beta_2)^\top = (0, 1, 0.8)^\top$, $n \in \{10, 20\}$, $\tau^2 \in \{0.1, 0.5, 1\}$, $\xi_i \sim N(0, 1)$ and $\xi_i \sim SN(0, 1, -5)$. In the last two scenarios, $\left(\mu_\zeta, \sigma_\zeta^2\right)^\top$ is set to $(0, 1)^\top$. This simulation follows three main steps, namely, generating the true measures of risk and covariate, generating outcomes, and computing the observed measures and the associated covariance matrices. In the last two scenarios, outcomes are generated at the group level and the subgroup level. In model (3.8), probabilities p_{iT1M} and p_{iC1M} are simulated from uniform distributions $U[0, \min\{\text{expit}(\eta_i), \text{expit}(\zeta_i)\}]$ and $U[0, \min\{\text{expit}(\xi_i), \text{expit}(\zeta_i)\}]$, respectively. In model (3.9), within-study correlations ρ_{iT} and ρ_{iC} are generated from a uniform distribution $U(-1, 1)$. The simulation set-up considers $1,000$ replicates for each scenario. The simulation is carried out using the R programming language (R Core Team, 2021).

Assuming the approximate measurement error model, model (3.1) is fitted to data from the first two scenarios using the likelihood approach, while the pseudo-likelihood approach is applied to fit model (3.2) to data from the last two scenarios. Furthermore, in the last two scenarios, the pseudo-likelihood approach is compared with the likelihood approach based on subgroup summary information. The performance of each of these methods is evaluated using the bias, the standard error (se) and the standard deviation (sd) of the estimator of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$, and the number of convergent solutions from that method (conv). Moreover, the empirical coverage probability is computed for 95% confidence intervals of $(\beta_0, \beta_1, \beta_2)^\top$ using the Hessian standard error and the sandwich standard error. Likelihood estimators and pseudo-likelihood estimators are also compared to the naive analysis ignoring the measurement error based on their empirical coverage probabilities.

3.3.2 Simulation results

Table 3.3 shows the bias, the standard errors and the standard deviations of the maximum likelihood estimators of all parameters in model (3.1) when data are from scenario 1. The likelihood approach underestimates β_2 and τ^2. However, the estimators of regression coefficients are satisfactory since they have very small bias. Furthermore, the relative bias of the estimator of τ^2 decreases in magnitude with the number of studies and the residual variance. The standard errors of the estimators of $(\beta_0, \beta_1, \beta_2, \tau^2)^\top$ decline when there are more studies. These standard errors are slightly different from

the associated standard deviations and the discrepancies between them become smaller when increasing n. Figure 3.1 shows that the empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2)^\top$ from the likelihood approach are below the nominal level and they approach this level when the number of studies grows. Furthermore, the confidence interval for β_1 from the likelihood approach has higher empirical coverage probability compared to the naive analysis, especially when the residual variance is small. The percentage of convergent solutions is satisfactory and it inclines with n and τ^2. Similar results can be found for data from the skewed control risk distribution or model (3.1) with $\zeta_i \sim \text{Bernoulli}\,(0.5)$ fixed. See Tables A.1, A.2 and Figures A.1, 3.2 and A.2 for further details.

Tables 3.5, A.8, and A.9 show the bias, the standard errors and the standard deviations of the likelihood estimators and the pseudo-likelihood estimators of all parameters in model (3.2) when data are from scenario 3. The pseudo-likelihood approach and the likelihood approach based on subgroup summary have the similar and satisfactory performance. Specifically, the estimators of the regression coefficients are nearly unbiased. Although the residual variance is underestimated, the relative bias of its estimator declines in magnitude with the number of studies and the residual variance. The standard errors of the estimators of $(\beta_0, \beta_1, \beta_2, \tau^2)^\top$ drop when increasing n. Moreover, the differences between the standard errors and the associated standard deviations are small and they reduce when increasing number of studies included in the meta-analysis. However, the difference between the standard error and the standard deviation of the estimator of β_1 is slightly large when n is small and τ^2 is sizable. The percentage of convergent solutions is approximately 100%. Figure 3.3 shows that the empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2)^\top$ from the pseudo-likelihood approach are high and they approach the nominal level when the number of studies increases. The confidence interval for β_1 from the pseudo-likelihood approach has higher empirical coverage probability compared to the naive analysis, especially when the residual variance is small. When data are from scenario 4, the results are similar, but the difference between the standard error and the standard deviation of the estimator of β_2 is slightly large if n is small and τ^2 is substantial. See Tables 3.6, A.3, and A.4 and Figure 3.4 for more details.

Tables A.5, A.6, and A.7 show the bias, the standard errors and the standard deviations of the likelihood estimators and the pseudo-likelihood estimators when data are from scenario 4 with $\xi_i \sim SN\,(0, 1, -5)$. Although the results are less satisfactory compared to scenarios of the normal control risk distribution, they tend to improve when the number of studies increases. Both the likelihood approach and the pseudo-likelihood approach

overestimate β_1 when n is small and τ^2 is moderate. Furthermore, there are considerable differences between the standard errors and the associated standard deviations of the estimator of β_0 and β_1 if the number of studies is small and the residual variance is moderate to large, as expected. Moreover, the empirical coverage probabilities of 95% confidence intervals for the regression coefficients from the likelihood approach and the pseudo-likelihood approach are under the nominal level and they approach this level when the number of studies grows. The confidence intervals for β_0 and β_1 from the likelihood approach and the pseudo-likelihood approach has higher empirical coverage probabilities compared to the naive analysis, especially when the residual variance is small. The percentage of convergent solutions is still nearly 100% and it increases with the number of studies and the residual variance. See Figure A.3. Similar results can be found for data from scenario 3 with $\xi_i \sim SN(0, 1, -5)$. See Tables A.10, A.11, and A.12 and Figure A.4 for more details. However, this scenario is slightly different from the previous scenario. Specifically, the likelihood approach and the pseudo-likelihood approach overestimate β_0 and β_1 when the number of studies is moderate and the residual variance is large. Furthermore, the pseudo-likelihood estimators of β_0 and β_1 have much smaller standard errors compared to the likelihood estimators when n and τ^2 are small.

TABLE 3.3: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach and the likelihood approach when data follows scenario 1. Underlying risk distributed as a standard normal.

τ^2	par.	appr.	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.1	β_0	lik	0.007	0.134	0.160	991	-0.006	0.096	0.102	998
		naive	0.019	0.149	0.176	1000	-0.003	0.104	0.121	1000
	β_1	lik	0.011	0.172	0.214	991	0.012	0.121	0.136	998
		naive	-0.032	0.164	0.227	1000	-0.061	0.107	0.147	1000
	β_2	lik	-0.014	0.134	0.160	991	-0.010	0.119	0.128	998
		naive	0.014	0.152	0.180	1000	0.014	0.140	0.157	1000
	μ_ξ	lik	-0.010	0.302	0.314	991	0.016	0.219	0.224	998
	τ^2	lik	-0.056	0.048	0.062	982	-0.036	0.050	0.055	996
		naive	0.091	0.072	0.125	1000	0.120	0.064	0.090	1000
	σ_ξ^2	lik	-0.088	0.440	0.447	991	-0.072	0.320	0.322	998
0.5	β_0	lik	-0.000	0.232	0.295	998	0.001	0.165	0.182	1000
		naive	-0.004	0.340	0.405	1000	-0.008	0.179	0.214	1000
	β_1	lik	0.019	0.282	0.363	998	0.012	0.195	0.213	1000
		naive	-0.027	0.293	0.376	1000	-0.042	0.187	0.218	1000
	β_2	lik	-0.023	0.195	0.247	998	-0.017	0.186	0.204	1000
		naive	0.014	0.280	0.342	1000	0.023	0.222	0.248	1000
	μ_ξ	lik	-0.002	0.299	0.319	998	-0.006	0.217	0.221	1000
	τ^2	lik	-0.217	0.188	0.233	994	-0.128	0.161	0.171	1000
		naive	0.114	0.230	0.403	1000	0.094	0.173	0.217	1000
	σ_ξ^2	lik	-0.106	0.434	0.462	998	-0.093	0.315	0.307	1000
1	β_0	lik	0.009	0.296	0.374	999	-0.011	0.251	0.273	1000
		naive	0.009	0.346	0.432	1000	-0.006	0.258	0.315	1000
	β_1	lik	-0.003	0.381	0.472	999	-0.007	0.254	0.275	1000
		naive	-0.046	0.408	0.474	1000	-0.058	0.248	0.296	1000
	β_2	lik	-0.012	0.434	0.520	999	-0.009	0.285	0.327	1000
		naive	0.024	0.455	0.576	1000	0.016	0.311	0.380	1000
	μ_ξ	lik	-0.005	0.293	0.328	999	-0.005	0.218	0.223	1000
	τ^2	lik	-0.365	0.358	0.411	996	-0.251	0.290	0.314	1000
		naive	-0.035	0.361	0.539	1000	0.066	0.311	0.397	1000
	σ_ξ^2	lik	-0.143	0.417	0.441	999	-0.088	0.317	0.318	1000

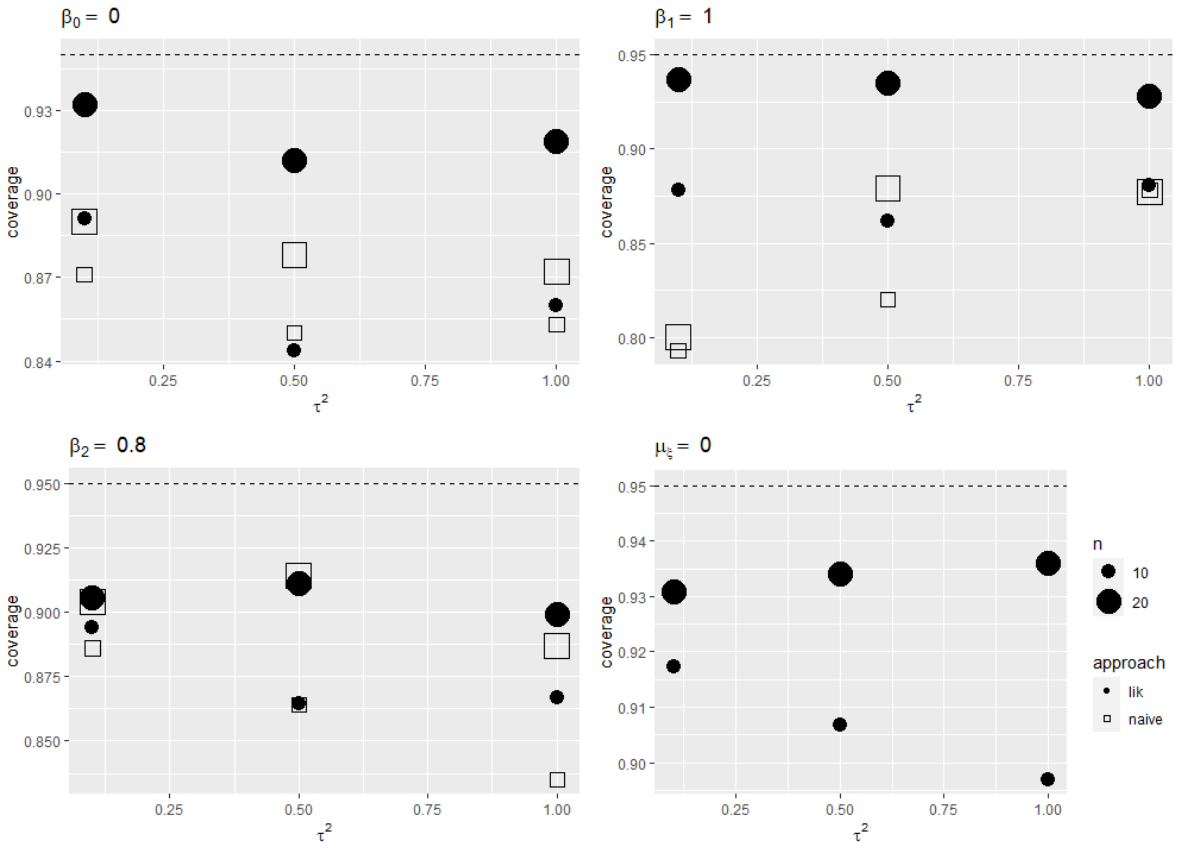

FIGURE 3.1: Empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi)^\top$ for the uncorrected approach and the likelihood approach when data follows scenario 1. Underlying risk distributed as a standard normal.

TABLE 3.4: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2\right)^\top$ and number of convergent solutions over 1,000 replicates for the uncorrected approach and the likelihood approach when data follows scenario 2. Underlying risk distributed as a standard normal.

τ^2	par.	appr.	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.1	β_0	lik	0.004	0.152	0.184	989	0.001	0.161	0.170	996
		naive	0.006	0.164	0.215	1000	0.004	0.184	0.199	1000
	β_1	lik	0.010	0.174	0.205	989	0.004	0.124	0.130	996
		naive	-0.048	0.164	0.215	1000	-0.052	0.111	0.138	1000
	β_2	lik	-0.025	0.285	0.349	989	-0.021	0.202	0.214	996
		naive	0.003	0.341	0.398	1000	0.012	0.228	0.246	1000
	μ_ξ	lik	-0.000	0.297	0.323	989	-0.007	0.220	0.221	996
	τ^2	lik	-0.055	0.049	0.060	983	-0.033	0.052	0.059	988
		naive	0.089	0.071	0.120	1000	0.148	0.072	0.105	1000
	σ_ξ^2	lik	-0.120	0.426	0.424	989	-0.068	0.323	0.317	996
0.5	β_0	lik	-0.003	0.294	0.356	999	-0.013	0.238	0.264	1000
		naive	0.006	0.320	0.390	1000	-0.014	0.232	0.291	1000
	β_1	lik	-0.003	0.281	0.328	999	0.014	0.201	0.211	1000
		naive	-0.051	0.289	0.337	1000	-0.049	0.189	0.223	1000
	β_2	lik	-0.031	0.425	0.498	999	-0.019	0.327	0.354	1000
		naive	0.010	0.512	0.595	1000	0.031	0.371	0.413	1000
	μ_ξ	lik	-0.001	0.297	0.319	999	-0.018	0.216	0.233	1000
	τ^2	lik	-0.204	0.188	0.226	998	-0.128	0.165	0.175	1000
		naive	0.042	0.203	0.317	1000	0.125	0.182	0.233	1000
	σ_ξ^2	lik	-0.121	0.427	0.431	999	-0.111	0.310	0.313	1000
1	β_0	lik	0.021	0.439	0.585	1000	-0.003	0.380	0.445	1000
		naive	0.019	0.539	0.631	1000	0.007	0.378	0.485	1000
	β_1	lik	0.024	0.363	0.464	1000	-0.013	0.262	0.290	1000
		naive	-0.014	0.387	0.467	1000	-0.063	0.253	0.319	1000
	β_2	lik	-0.058	0.568	0.740	1000	-0.046	0.461	0.538	1000
		naive	-0.008	0.697	0.828	1000	0.000	0.489	0.614	1000
	μ_ξ	lik	0.007	0.295	0.325	1000	0.003	0.216	0.214	1000
	τ^2	lik	-0.418	0.335	0.403	1000	-0.267	0.292	0.296	1000
		naive	-0.111	0.333	0.522	1000	0.119	0.326	0.400	1000
	σ_ξ^2	lik	-0.137	0.420	0.439	1000	-0.112	0.311	0.319	1000

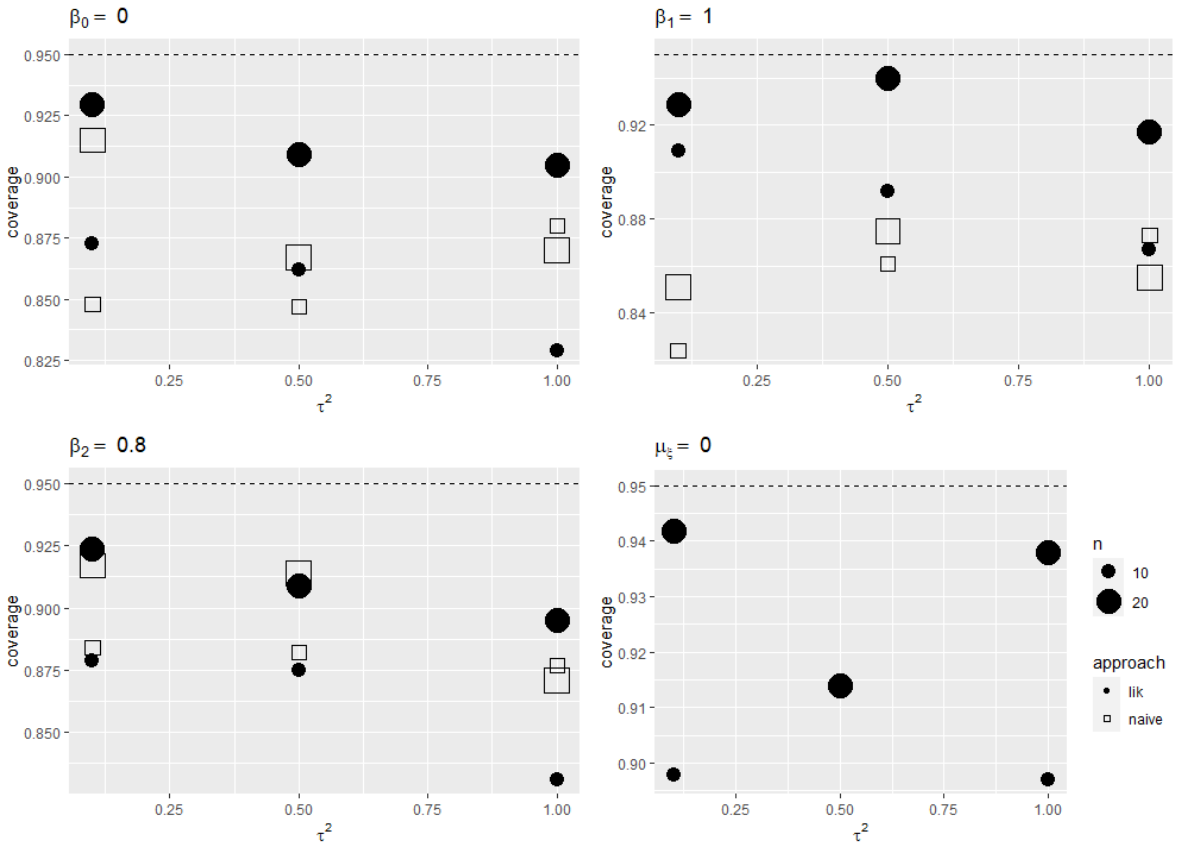

FIGURE 3.2: Empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi)^\top$ for the uncorrected approach and the likelihood approach when data follows scenario 2. Underlying risk distributed as a standard normal.

TABLE 3.5: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and number of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 3 and $\tau^2 = 0.1$. Underlying risk distributed as a standard normal.

τ^2	par.	approach	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.1	β_0	lik	0.000	0.150	0.166	992	0.007	0.102	0.109	997
		pseudo-likelihood	-0.001	0.151	0.167	988	0.007	0.104	0.109	994
		naive	-0.002	0.161	0.184	1000	0.008	0.112	0.125	1000
	β_1	lik	0.004	0.185	0.203	992	0.009	0.126	0.131	997
		pseudo-likelihood	0.006	0.186	0.206	988	0.008	0.126	0.130	994
		naive	-0.036	0.169	0.211	1000	-0.064	0.111	0.139	1000
	β_2	lik	-0.012	0.168	0.197	991	-0.003	0.113	0.124	997
		pseudo-likelihood	-0.012	0.171	0.197	988	-0.001	0.115	0.125	994
		naive	-0.015	0.171	0.211	1000	-0.002	0.116	0.139	1000
	μ_ξ	lik	0.008	0.302	0.309	993	-0.001	0.217	0.225	997
		pseudo-likelihood	0.012	0.303	0.309	988	-0.003	0.219	0.229	994
	μ_ζ	lik	0.002	0.299	0.315	993	0.000	0.215	0.222	997
		pseudo-likelihood	0.004	0.300	0.319	988	0.001	0.216	0.224	994
	τ^2	lik	-0.046	0.060	0.066	979	-0.030	0.054	0.056	987
		pseudo-likelihood	-0.047	0.059	0.066	978	-0.031	0.054	0.058	984
		naive	0.099	0.075	0.124	1000	0.147	0.072	0.107	1000
	σ_ξ^2	lik	-0.090	0.444	0.451	993	-0.089	0.316	0.318	997
		pseudo-likelihood	-0.083	0.446	0.447	988	-0.079	0.320	0.320	994
	σ_ζ^2	lik	-0.087	0.428	0.423	993	-0.079	0.305	0.306	997
		pseudo-likelihood	-0.079	0.431	0.424	988	-0.069	0.308	0.308	994

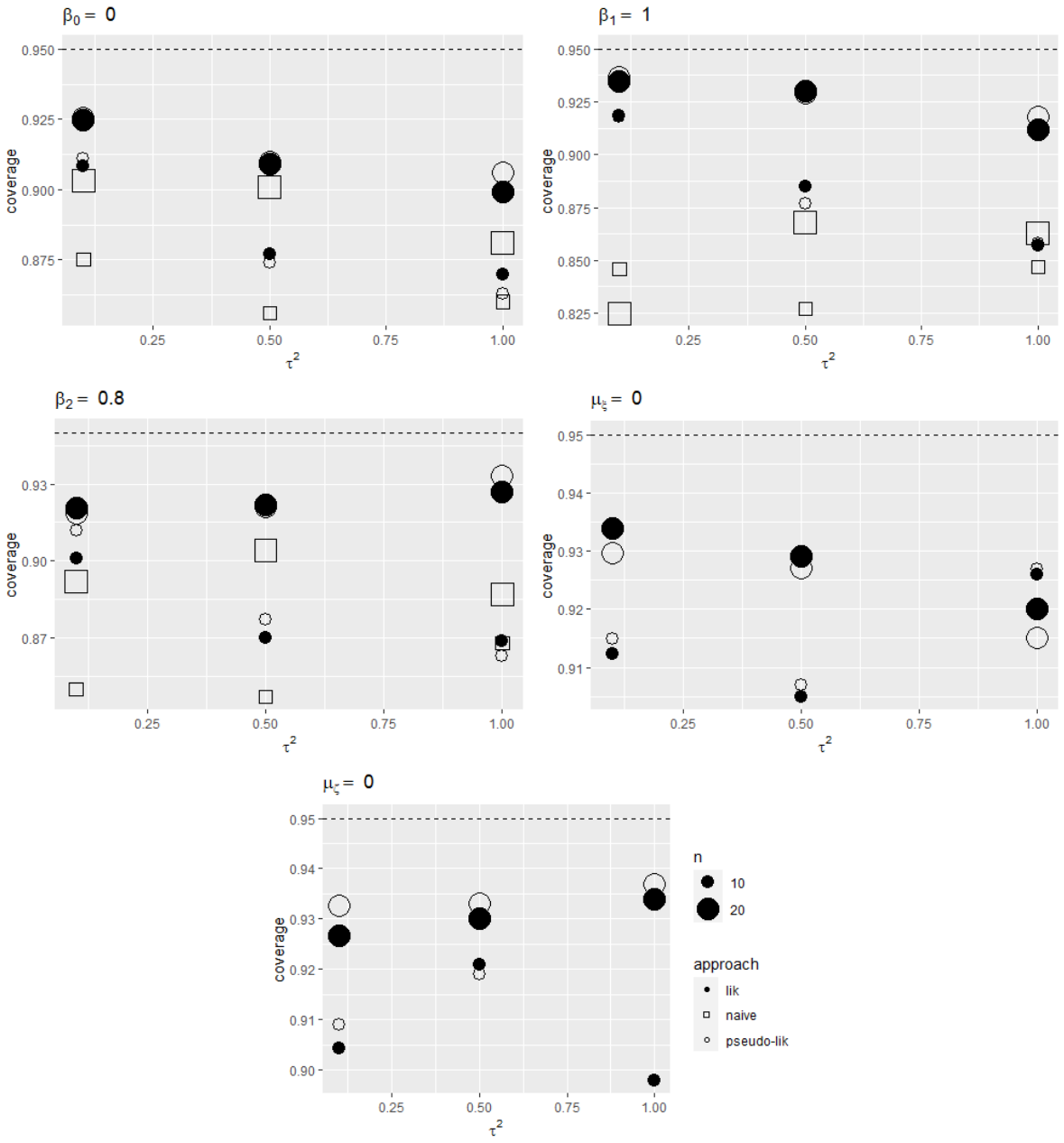

FIGURE 3.3: Empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta)^\top$ for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 3. Underlying risk distributed as a standard normal.

TABLE 3.6: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma^2_\xi, \sigma^2_\zeta\right)^\top$ and number of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 4 and $\tau^2 = 0.1$. Underlying risk distributed as a standard normal.

τ^2	par.	approach	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.1	β_0	lik	-0.004	0.148	0.170	995	0.001	0.104	0.118	996
		pseudo-lik	-0.004	0.148	0.169	992	0.001	0.104	0.118	994
		naive	-0.005	0.160	0.185	1000	0.003	0.114	0.134	1000
	β_1	lik	0.008	0.185	0.216	995	0.006	0.129	0.143	996
		pseudo-lik	0.009	0.186	0.216	992	0.007	0.129	0.144	994
		naive	-0.047	0.170	0.225	1000	-0.084	0.111	0.148	1000
	β_2	lik	-0.001	0.165	0.191	995	-0.012	0.111	0.124	996
		pseudo-lik	-0.002	0.165	0.192	992	-0.013	0.111	0.123	994
		naive	0.017	0.176	0.225	1000	0.011	0.120	0.144	1000
	μ_ξ	lik	0.005	0.299	0.326	995	-0.001	0.220	0.229	996
		pseudo-lik	0.005	0.299	0.323	992	-0.001	0.220	0.228	994
	μ_ζ	lik	0.008	0.293	0.322	995	0.003	0.216	0.217	996
		pseudo-lik	0.006	0.294	0.325	992	0.002	0.216	0.215	994
	τ^2	lik	-0.045	0.059	0.069	985	-0.033	0.054	0.056	984
		pseudo-lik	-0.046	0.059	0.069	987	-0.034	0.054	0.056	981
		naive	0.102	0.075	0.135	1000	0.147	0.072	0.100	1000
	σ^2_ξ	lik	-0.109	0.436	0.441	995	-0.080	0.325	0.311	996
		pseudo-lik	-0.107	0.437	0.450	992	-0.080	0.325	0.316	994
	σ^2_ζ	lik	-0.093	0.411	0.445	995	-0.049	0.304	0.308	996
		pseudo-lik	-0.092	0.411	0.441	992	-0.051	0.303	0.306	994

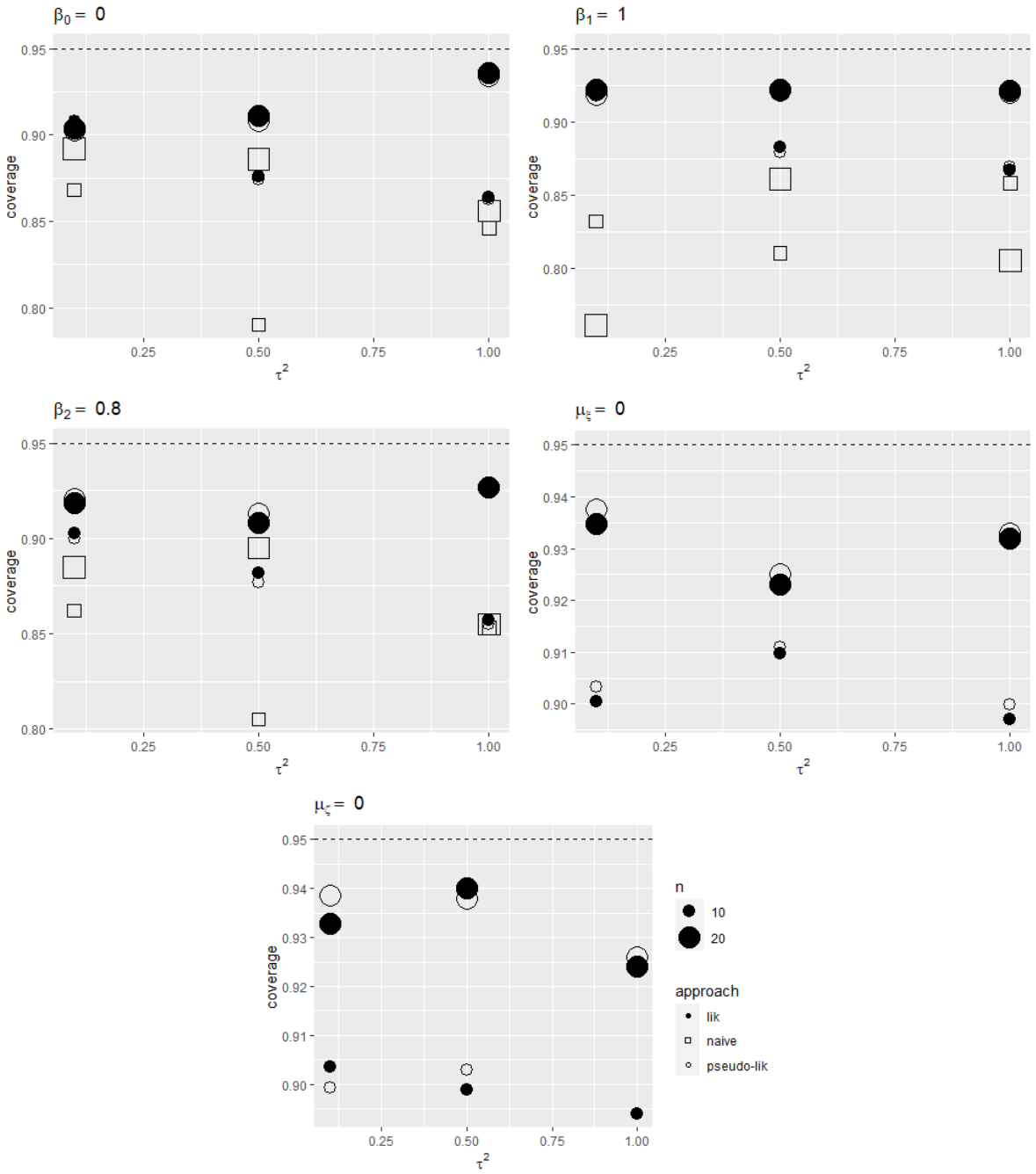

FIGURE 3.4: Empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta)^\top$ for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 4. Underlying risk distributed as a standard normal.

3.3.3 A simulation study for the exact pseudo-likelihood approach

The previous simulation study has been focused on the approximate measurement error model, mainly for computational convenience. Under an exact measurement error model, in fact, the likelihood function is very complex, as shown in Section 3.2.4.
In order to investigate the behavior of the pseudo-likelihood approach under an exact measurement error model, a small simulation study has been performed. The simulation study refers to data from scenario 3 with $(n, \tau^2)^\top = (10, 0.5)^\top$. When evaluating the pseudo-likelihood function using Gaussian-Hermite quadrature, the number of nodes in each dimension is set equal to 10. Only 100 replicates of the simulation are considered, for computational reasons, since the estimation process is time consuming, as it takes more than 10 hours to run with HPE SuperDome Flex 280 server. Results are reported in Table 3.7.
These results confirm the previous findings under the approximate measurement error model. The pseudo-likelihood approach provides estimators of the parameters with small bias, especially when focusing on regression coefficients. The slight differences between standard errors and the associated standard deviations are expected since we are adopting a pseudo-likelihood solution under a working independence assumption. The empirical coverage probabilities associated to the regression coefficients estimators from all the measurement error models are lower than the nominal level. Furthermore, the approximate approach is preferable in this simulation because of its better performance. The results from the exact approach may be affected by the small number of nodes as well as the small sample size.

3.4 Examples

3.4.1 Schizophrenia dataset

Populations with older age, unhealthy lifestyle, physical comorbidities and psychiatric diseases are easily affected by COVID-19 (Gold *et al.*, 2020, Hariyanto and Kurniawan, 2021, Pardamean *et al.*, 2022). There exist risk factors in schizophrenic patients which are known to increase the risk of getting worse impacts of COVID-19 (Xiong *et al.*, 2020). For example, there exists a dysregulated immune response in schizophrenic patients which can increase the risk of mortality due to COVID-19 (Rodrigues-Amorim *et al.*, 2018, Kroken *et al.*, 2019). Therefore, Pardamean *et al.* (2022) conducted a meta-analysis of 10 studies to investigate the relationship between schizophrenia and mortality

TABLE 3.7: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$, empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta)^\top$ and number of convergent solutions over 100 replicates for the pseudo-likelihood approach. Data follows scenario 3 and $\left(n, \tau^2\right)^\top = (10, 0.5)^\top$. Underlying risk distributed as a standard normal.

par.	m.e. model	bias	se	sd	emp. coverage prob.	conv.
β_0	exact	-0.009	0.233	0.262	0.82	100
	approx.	-0.011	0.238	0.249	0.88	100
β_1	exact	-0.010	0.266	0.319	0.83	100
	approx.	-0.016	0.280	0.314	0.88	100
β_2	exact	0.002	0.264	0.360	0.81	100
	approx.	-0.016	0.271	0.354	0.87	100
μ_ξ	exact	0.020	0.309	0.312	0.92	100
	approx.	0.027	0.300	0.300	0.94	100
μ_ζ	exact	-0.084	0.300	0.322	0.89	100
	approx.	-0.087	0.296	0.319	0.90	100
τ^2	exact	-0.178	0.205	0.243	-	100
	approx.	-0.196	0.200	0.227		98
σ_ξ^2	exact	-0.041	0.474	0.528	-	100
	approx.	-0.103	0.440	0.455		100
σ_ζ^2	exact	-0.058	0.444	0.519	-	100
	approx.	-0.091	0.424	0.477		100

due to COVID-19. Their dataset is shown in Table 3.8, including the numbers of deaths and the sizes of two groups, namely, a group with schizophrenia and a group without schizophrenia. The mean age, the percentage of male patients, and the percentage of diabetic patients are also reported in their meta-analysis. Figure 3.5 shows a forest plot of this dataset, highlighting a substantial heterogeneity among studies since the variability of effects expressed by the length of confidence intervals is evident. See, for example, the heterogeneity between the effect estimates in Fond *et al.* (2020) and Rivas-Ramírez *et al.* (2021). Pardamean *et al.* (2022) found that schizophrenic patients have a higher risk of mortality due to COVID-19 compared to patients without schizophrenia ($RR = 2.22$; 95%CI: $(1.54, 3.20)$). Since the heterogeneity between studies is large ($I^2 = 82\%$), they fit meta-regression models with several risk factors, namely, age, gender, hypertension, diabetes, smoking, obesity, and mood disorders, to the dataset using the restricted likelihood approach. Associations between mortality in schizophrenic patients and mean age $\left(\hat{\beta}_{\text{age}} = -0.0334\text{; 95\%CI: } (-0.0519, -0.0150)\right)$ and the percentage of smokers $\left(\hat{\beta}_{\text{smoker}} = 0.0269\text{; 95\%CI: } (0.0082, 0.0456)\right)$ are statistically significant.
Let the log odds of mortality due to COVID-19 be the risk measure. In this section,

we want to examine associations between the risk of mortality in schizophrenic patients and the risk of mortality in patients without schizophrenia, the mean age, the log odds of male and the log odds of diabetes. These associations are graphically shown in Figure 3.6. The scatter plots suggest linear relationships between the risk measures between the two groups of patients, the mean age, the log odds of diabetes. A control risk regression model (2.1) is fitted to the dataset using the likelihood approach. Furthermore, control risk regression models with each of the covariates of interest are also fitted to the dataset. The mean age is scaled or standardized in order to avoid computational issues. Other study-specific characteristics such as the percentage of hypertension are discarded because they have too many missing values. Since information about within-study covariances between the additional covariates and the risk measures is unavailable, the pseudo-likelihood approach developed in this chapter is applied when fitting model (3.2). Results of the likelihood approach and the pseudo-likelihood approach are compared with the results of the naive analysis.

Table 3.9 shows the estimates of regression coefficients and the residual variance from the likelihood approach, the pseudo-likelihood approach and the naive analysis. According to all the approaches, the coefficient of ξ_i is positive and statistically different from zero in model (2.1) and model (3.2). Therefore, there exists a significant association between the risk measures in the group of patients with the pathology and the group of patients without the pathology. In the classical model, β_1 is smaller than 1, so there is an indication of reduced risk of mortality for pathological patients. In the model with the (scaled) mean age, the naive estimate and the pseudo-likelihood estimate of β_1 are not significantly different from one, while the pseudo-likelihood finds a significant association between the risk of mortality in the pathological group with the covariate. As a consequence, the residual variance is very small. Therefore, given the results for the model with an additional covariate given by age, the pseudo-likelihood approach suggests that the difference between the risk of mortality for pathological patients and patients in the control condition tends to decrease with age. However, these results are based on a small number of studies. When inserting covariates results are similar using pseudo-likelihood approach in place of the likelihood approach for log odds of male and log odds of diabetes. Specifically, their estimates of β_1 are smaller than one. Inserting an additional covariate and accounting for the error in the summarized data reduce the amount of the residual variance with respect ot the naive analysis.

With schizophrenia		Without schizophrenia		Characteristics		
Events	Total	Events	Total	Mean age	% Male	% Diabetes
20	984	632	127281	40.3 ± 20.8	48.4	4.6
22	649	7	709	51.5 ± 15.4	60.9	17.1
4	15	94	1077	63.1 ± 18.5	54.3	23.4
211	823	10854	49927	70.3 ± 19.2	56.8	27.8
6	159	49	2817	55.4 ± 16.2	41.7	15.1
20	75	701	6349	54 ± 18.6	47	25.7
11	40	760	4372	67.7 ± 20.7	41.2	11.9
2	18	3	69	51.5 ± 14.8	47.1	4.5
2	4	77	414	65.4 ± 16.6	56.9	23.6
5	6	70	144	77.6 ± 10.5	50	34

TABLE 3.8: Schizophrenia dataset (Pardamean *et al.*, 2022).

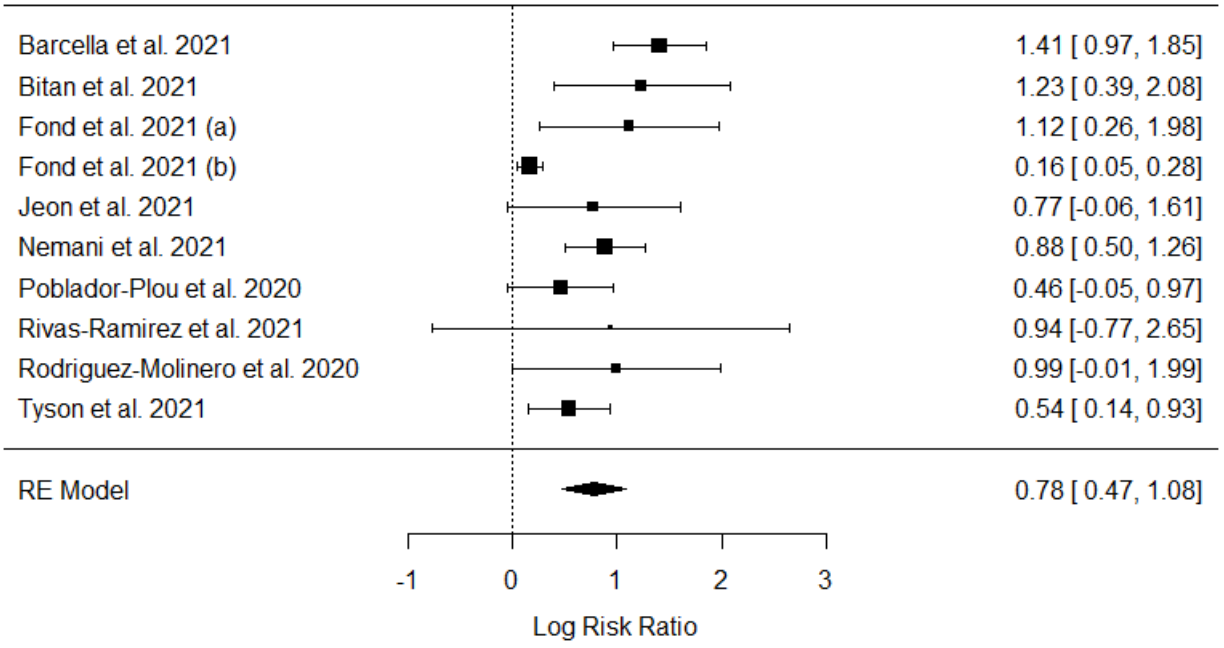

FIGURE 3.5: Forest plot for schizophrenia dataset (Pardamean *et al.*, 2022).

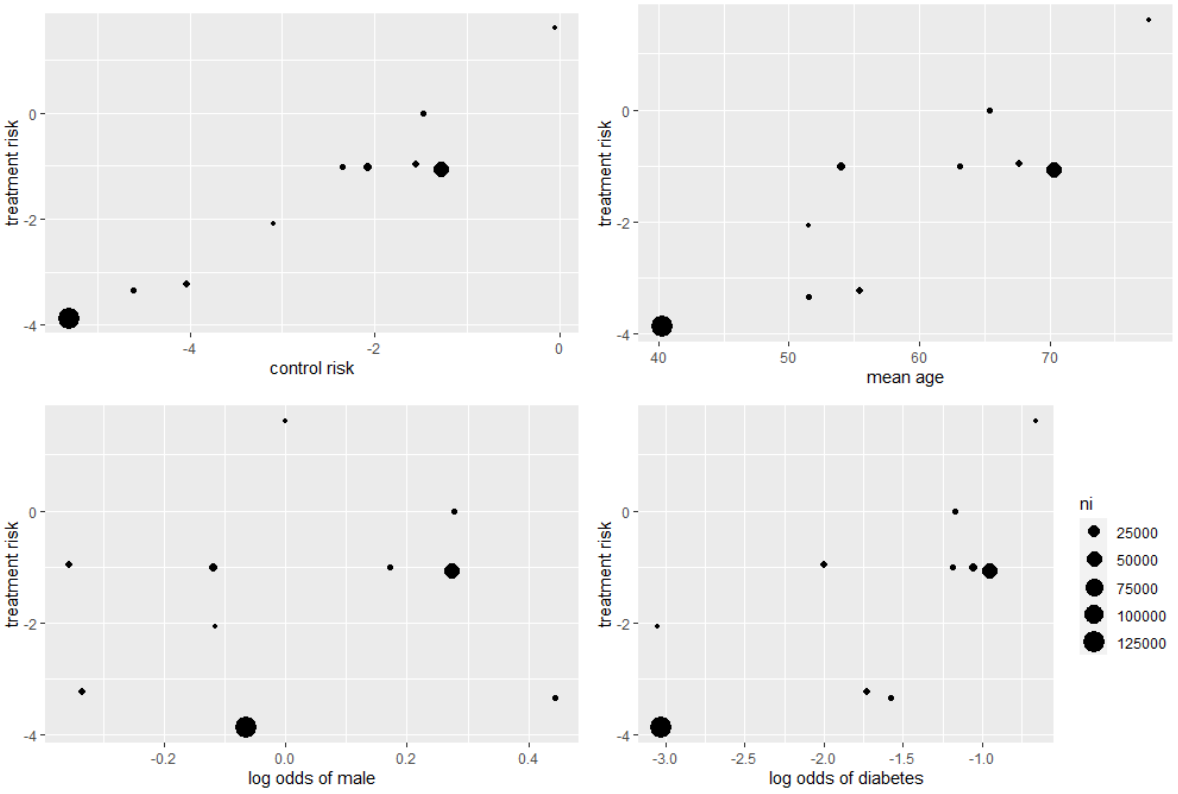

FIGURE 3.6: Scatter plots of the risk measures in pathological group and non-pathological group, the mean age, the log odds of male and the log odds of diabetes. Schizophrenia dataset (Pardamean *et al.*, 2022).

Covariate	Approach	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}^2$
No	likelihood	0.146 (0.329)	**0.761 (0.091)**	-	0.028 (0.049)
	naive analysis	-0.080 (0.140)	**0.709 (0.056)**	-	0.527 (0.211)
Scaled mean age	pseudo-likelihood	0.903 (0.486)	**1.079 (0.186)**	**-0.570 (0.270)**	2.5e-05 (0.001)
	naive analysis	0.740 (0.410)	**1.004 (0.149)**	**-0.492 (0.236)**	0.466 (0.174)
Log odds of male	pseudo-likelihood	0.121 (0.186)	**0.755 (0.063)**	-0.612 (0.418)	0.002 (0.008)
	naive analysis	0.034 (0.177)	**0.724 (0.057)**	-0.422 (0.407)	0.492 (0.184)
Log odds of diabetes	pseudo-likelihood	0.138 (0.585)	**0.766 (0.137)**	-0.014 (0.269)	0.028 (0.061)
	naive analysis	-0.092 (0.195)	**0.718 (0.110)**	-0.025 (0.251)	0.524 (0.196)

TABLE 3.9: Estimates of $\left(\beta_0, \beta_1, \beta_2, \tau^2\right)^\top$ and the associated standard errors when using the likelihood approach to fit the classical control risk regression model and the pseudo-likelihood approach to fit the control risk regression model with one additional covariate to schizophrenia dataset (Pardamean *et al.*, 2022). Significant coefficients are highlighted.

3.4.2 Myocardial injury dataset

Nowadays, data are needed to well stratify COVID-19 patients based on clinical and laboratory parameters to avoid serious COVID-19 outcomes (Sanz-Sánchez *et al.*, 2021). Myocardial involvements have been observed in COVID-19 patients and associated with bad outcomes (see, e.g., Guo *et al.*, 2020, Shi *et al.*, 2020). Therefore, Sanz-Sánchez *et al.* (2021) conducted a meta-analysis of 14 studies to evaluate the impact of myocardial injury (MI) on all-cause mortality in COVID-19 patients. The dataset is shown in Table 3.10, including the numbers of deaths and the sizes of the group of patients with MI and the group without MI. The mean age, the percentage of males, the percentage of diabetic patients, and the percentage of patients with hypertension are also reported. A forest plot of the dataset is shown in Figure 3.7 and suggests the large heterogeneity across studies. (Sanz-Sánchez *et al.*, 2021) found that COVID-19 patients with MI have a higher risk of all-cause mortality compared to COVID-19 patients without MI ($OR = 9.16$; 95%CI: $(5.30, 15.83)$).

Figure 3.8 shows scatter plots of risk measures, the mean age, the log odds of male and the log odds of hypertension, suggesting linear relationships between the risk of mortality for pathological patients and non-pathological patients, the mean age, and the log odds of hypertension. In this section, we want to fit model (2.1) and model (3.2) to the dataset where the log odds of mortality is chosen as the risk measure. Consider the mean age, the log odds of male, and log odds of hypertension as additional covariates in control risk regression. The mean age is scaled or standardized in order to avoid computational issues. While the classical model is fitted with the likelihood approach, models with

additional covariates are fitted with the pseudo-likelihood approach. However, since we have no information about the variances of the ages of patients in this meta-analysis, we set the within-study variance of the mean ages to zero, which means that we assume no variation within study for this covariate. We also remove studies whose values of the additional covariates are missing.

Table 3.11 shows the estimates of the regression coefficients, the residual variance, and their associated standard errors from the likelihood approach, the pseudo-likelihood approach and the naive analysis. The estimates of the coefficient of the risk in the control condition from all the approaches are much smaller than one. Furthermore, this coefficient is significantly different from one, which is an indication of reduced risk of mortality for patients with MI. However, the relationship between the risk measures in the classical model is not statistically significant. According to the pseudo-likelihood approach, β_1 is significantly different from zero only in models with the scaled mean age and the log odds of hypertension. Furthermore, the pseudo-likelihood approach results in the negative and statistically significant coefficient of the log odds of hypertension. In other words, the difference between the risk of mortality for pathological patients and the risk of mortality for patients in the control conditions tends to decrease with the risk of hypertension. This might be due to the fact that patients with hypertension already followed therapies which take into account MI or COVID-19. Finally, the naive estimate of the residual variance is much larger than the pseudo-likelihood estimate, which shows that inserting an additional covariate and accounting for the error in the summarized data reduce the amount of the residual variance with respect ot the naive analysis.

TABLE 3.10: Myocardial injury dataset (Sanz-Sánchez *et al.*, 2021).

With MI		Without MI		Characteristics			
Events	Total	Events	Total	Age	% Male	% Diabetes	% Hypertension
68	94	15	109	62	62	17	34
13	21	28	158	58	54	18	32
50	123	18	209	67	71	21	54
12	23	16	105	64	57	NA	NA
31	52	12	135	59	49	15	33
14	24	8	30	68	NA	15	24
46	112	26	112	67	57	25	75
504	914	302	1906	59	54	18	25
48	89	15	35	68	69	20	50
51	133	11	538	63	48	15	30
121	170	65	989	NA	NA	NA	NA
0	10	1	125	47	53	9	10
3	16	0	85	49	54	14	21
23	50	1	95	56	62	19	30

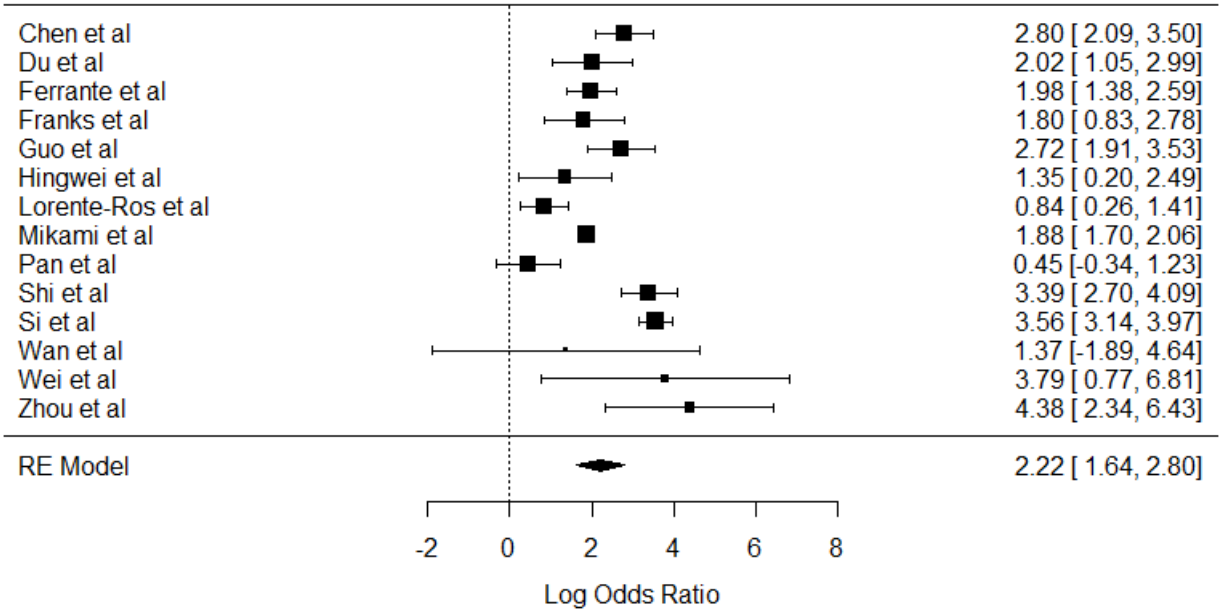

FIGURE 3.7: Forest plot for myocardial injury dataset (Sanz-Sánchez *et al.*, 2021).

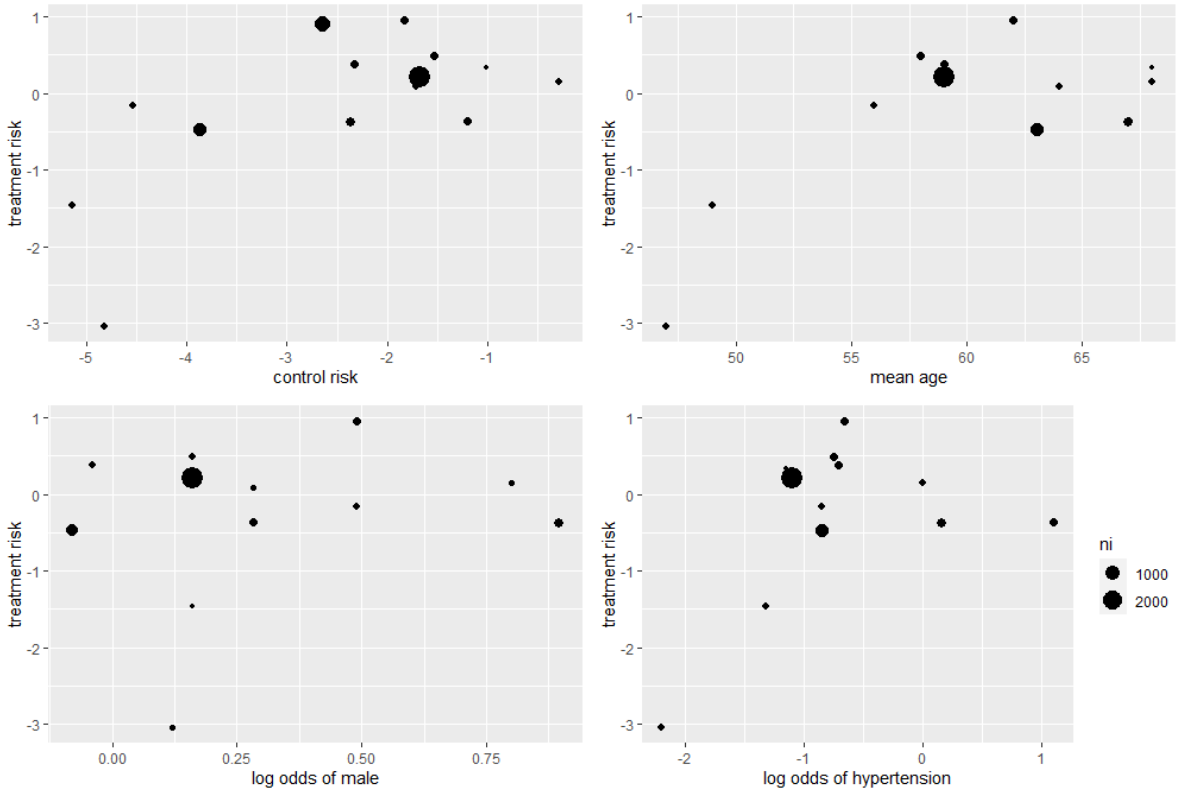

FIGURE 3.8: Scatter plot of the risk measures in pathological group and non-pathological group, the mean age, the log odds of male and the log odds of hypertension. Myocardial injury dataset (Sanz-Sánchez *et al.*, 2021).

TABLE 3.11: Estimates of $\left(\beta_0, \beta_1, \beta_2, \tau^2\right)^\top$ and their standard errors when using the likelihood approach and the naive analysis to fit the control risk regression model. Myocardial injury dataset (Sanz-Sánchez *et al.*, 2021).

Covariate	Approach	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}^2$
No	likelihood	0.568 (0.357)	0.230 (0.153)	-	0.203 (0.109)
	naive analysis	0.424 (0.275)	0.137 (0.127)	-	0.829 (0.290)
Scaled mean age	pseudo-likelihood	0.857 (0.390)	**0.376 (0.160)**	-0.143 (0.261)	0.114 (0.126)
	naive analysis	0.594 (0.212)	**0.251 (0.099)**	-0.300 (0.169)	0.921 (0.317)
Log odds of male	pseudo-likelihood	0.714 (0.445)	0.293 (0.150)	-0.320 (0.586)	0.118 (0.102)
	naive analysis	0.577 (0.287)	0.219 (0.116)	-0.271 (0.403)	0.737 (0.261)
Log odds of hypertension	pseudo-likelihood	0.533 (0.302)	**0.332 (0.111)**	**-0.356 (0.151)**	0.038 (0.082)
	naive analysis	0.338 (0.209)	**0.241 (0.095)**	**-0.292 (0.138)**	0.858 (0.303)

Chapter 4

Quadratic relationship between risk measures

The relationship between the control risk and the treatment risk is usually assumed linear mainly due to empirical consideration and for computational convenience. As an alternative, Arends *et al.* (2000) briefly mention the quadratic relationship between the risk measures without reasoning in their discussion section. More sophisticatedly, Boissel *et al.* (2008) simulate risk measures from models corresponding to the identified modes of action and discover that the linear relationship only holds in a short range of frequency of event. Because the number of studies is limited, the linear relationship shows an incomplete understanding of the true relationship between the measures of risk. The true relationship is usually more complex than the linear one, and considers

$$\eta_i = f\left(\xi_i, \zeta_i\right),$$

where f denotes a general function whose form depends on the characteristics of participants, the disease and the outcome under consideration, and ζ_i denotes a vector of characteristics which are not accounted by the baseline risk (Boissel *et al.*, 2008). As an example, Wang *et al.* (2009) propose a model which accounts for a good effect and a bad effect of treatments based on a logistic model and a sigmoidal Emax model. This model characterized a U-shaped relationship between the absolute risk difference and the frequency of event in the control group, which means patients with moderate risk of disease benefits the most from the treatment and the effect is small for those with too low or too high risk. However, both Boissel *et al.* (2008) and Wang *et al.* (2009) choose the event frequency as their measure of risk. Furthermore, the model proposed in Wang *et al.* (2009) becomes a linear control risk regression model if converting the risk measures to log odds. Therefore, it is necessary to derive non-linear control risk

regression models for various types of risk measure.

4.1 Quadratic control risk regression model

Consider negative events as outcomes. Consider model (2.1), where the control risk measure can be centered for the ease of interpretation and to avoid numerical issues of estimation, namely,

$$\eta_i = \beta_0 + \beta_1 (\xi_i - \mu_\xi) + \varepsilon_i, \quad \xi_i \sim N\left(\mu_\xi, \sigma_\xi^2\right), \quad \varepsilon_i \sim N\left(0, \tau^2\right). \tag{4.1}$$

Inspired by Arends *et al.* (2000), in order to account for a potential nonlinear relationship between the treatment risk and control risk, we propose to extend model (4.1) to a quadratic control risk regression model,

$$\eta_i = \beta_0 + \beta_1 (\xi_i - \mu_\xi) + \beta_2 (\xi_i - \mu_\xi)^2 + \varepsilon_i, \quad \xi_i \sim N\left(\mu_\xi, \sigma_\xi^2\right), \quad \varepsilon_i \sim N\left(0, \tau^2\right), \tag{4.2}$$

where the vector of parameters of interest is $\theta = \left(\beta_0, \beta_1, \beta_2, \mu_\xi, \tau^2, \sigma_\xi^2\right)^\top$. The quadratic term in model (4.2) describes a U-shaped relationship between the two measures of risk and hence useful for treatments having a positive effect and a negative effect. A negative value shows that the treatment risk increases until the control risk reaches a certain point and then decreases. Conversely, a positive value of β_2 shows the opposite behavior. Model (4.2) reduces to the classical linear control risk regression model when β_2 is zero. Polynomials of degree higher than two, polynomials of non-integer degrees, spline models, and the model suggested in Wang *et al.* (2009) can also result in U-shaped relationships. However, their complex forms might reduce their applicability: the first two families are not so common in practice because of their complexity in computation and explanation, the spline family requires the choice of the location and the number of knots, and the model suggested in Wang *et al.* (2009) can not be used for types of risk measures which are defined on the real line. Furthermore, it is worth noting that too complicated models may suffer from overfitting and that standard approaches for inference, e.g., the likelihood approach, can not be a proper solution, when the number of studies becomes small.

4.2 Likelihood function

Although ξ_i^2 can be considered as an additional covariate, the likelihood approach and the pseudo-likelihood approach derived in the previous chapter cannot be used to

perform inference on θ, since ξ_i^2 is not normally distributed. However, applying the same technique as in case of the linear control risk regression model, a likelihood function in the quadratic model can be derived by marginalizing the joint distribution of the true measures of risk and the observed ones

$$L(\theta) \propto \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\hat{\eta}_i|\eta_i} f_{\hat{\xi}_i|\xi_i} f_{\eta_i|\xi_i} f_{\xi_i} d\eta_i d\xi_i,$$

or the joint distribution of the true measures of risk and the observed ones or the outcomes

$$L(\theta) \propto \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{Y_i|\eta_i} f_{X_i|\xi_i} f_{\eta_i|\xi_i} f_{\xi_i} d\eta_i d\xi_i,$$

where the measurement error models in the previous chapters remains unchanged. The only difference or challenge here is from $f_{\eta_i|\xi_i}$, as the true measures of treatment risk and control risk are not jointly normally distributed. Nevertheless, given ξ_i and the normality of ε_i, the distribution of η_i is still normal. This is true even if the model for true risk measures is more highly nonlinear or the baseline risk deviates from the normal distribution. Formally,

$$\eta_i|\xi_i \sim N\left\{\beta_0 + \beta_1(\xi_i - \mu_\xi) + \beta_2(\xi_i - \mu_\xi)^2, \tau^2\right\}.$$

Without loss of generality, consider a meta-analysis of dichotomous outcomes. A likelihood function for θ under the exact measurement model (2.4) is

$$L(\theta) \propto \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{\text{expit}(\eta_i)\}^{Y_i} \{1 - \text{expit}(\eta_i)\}^{n_{iT} - Y_i} \{\text{expit}(\xi_i)\}^{X_i} \{1 - \text{expit}(\xi_i)\}^{n_{iC} - X_i} \tag{4.3}$$

$$\times \frac{1}{\sqrt{\tau^2}} \phi\left\{\frac{\eta_i - \beta_0 - \beta_1(\xi_i - \mu_\xi) - \beta_2(\xi_i - \mu_\xi)^2}{\tau}\right\} \times \frac{1}{\sqrt{\sigma_\xi^2}} \phi\left(\frac{\xi_i - \mu_\xi}{\sigma_\xi}\right) d\eta_i d\xi_i,$$

and a likelihood function under the approximate measurement error model is

$$
\begin{aligned}
L(\theta) \propto & \prod_{i=1}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi\left(\frac{\hat{\eta}_i - \eta_i}{s_{\eta_i}}\right) \phi\left(\frac{\hat{\xi}_i - \xi_i}{s_{\xi_i}}\right) \\
& \times \frac{1}{\sqrt{\tau^2}} \phi\left\{\frac{\eta_i - \beta_0 - \beta_1\left(\xi_i - \mu_\xi\right) - \beta_2\left(\xi_i - \mu_\xi\right)^2}{\tau}\right\} \times \frac{1}{\sqrt{\sigma_\xi^2}} \phi\left(\frac{\xi_i - \mu_\xi}{\sigma_\xi}\right) \\
\approx & \prod_{i=1}^{n} \sum_{j=1}^{N} \sum_{k=1}^{N} \omega_j \omega_k g_i(x_j, y_k; \theta).
\end{aligned} \tag{4.4}
$$

The likelihood function does not have any closed-form expression, so numerical integration is needed. For instance, using a Gaussian-Hermite quadrature, the likelihood function can be approximated as

$$
L(\theta) \propto \prod_{i=1}^{n} \sum_{j=1}^{N} \sum_{k=1}^{N} \omega_j \omega_k g_i\left(u_j, u_k; \theta\right), \tag{4.5}
$$

with

$$
\begin{aligned}
g_i\left(u_j, u_k; \theta\right) = & \left\{\operatorname{expit}\left(u_j\right)\right\}^{Y_i} \left\{1 - \operatorname{expit}\left(u_j\right)\right\}^{n_{iT} - Y_i} \left\{\operatorname{expit}\left(u_k\right)\right\}^{X_i} \left\{1 - \operatorname{expit}\left(u_k\right)\right\}^{n_{iC} - X_i} \\
& \times \frac{1}{\sqrt{\tau^2}} \phi\left\{\frac{u_j - \beta_0 - \beta_1\left(u_k - \mu_\xi\right) - \beta_2\left(u_k - \mu_\xi\right)^2}{\tau}\right\} \times \frac{1}{\sqrt{\sigma_\xi^2}} \phi\left(\frac{u_k - \mu_\xi}{\sigma_\xi}\right) \\
& \times \exp\left(u_j^2\right) \exp\left(u_k^2\right),
\end{aligned}
$$

in (4.3) and

$$
\begin{aligned}
g_i\left(u_j, u_k; \theta\right) = & \phi\left(\frac{\hat{\eta}_i - u_j}{s_{\eta_i}}\right) \phi\left(\frac{\hat{\xi}_i - u_k}{s_{\xi_i}}\right) \\
& \times \frac{1}{\sqrt{\tau^2}} \phi\left\{\frac{u_j - \beta_0 - \beta_1\left(u_k - \mu_\xi\right) - \beta_2\left(u_k - \mu_\xi\right)^2}{\tau}\right\} \times \frac{1}{\sqrt{\sigma_\xi^2}} \phi\left(\frac{u_k - \mu_\xi}{\sigma_\xi}\right) \\
& \times \exp\left(u_j^2\right) \exp\left(u_k^2\right),
\end{aligned}
$$

in (4.4), where N denotes the number of nodes in each dimension of the integration, u_j and u_k denote the roots of the N-order Hermite polynomial, ω_j and $\omega_k, j, k = 1, \cdots, N$ denote the associated weights

$$
\omega_j = \frac{2^{N-1} N! \sqrt{\pi}}{N^2 \left\{H_{N-1}\left(u_j\right)\right\}},
$$

where H_{N-1} denotes the $N-1$-order Hermite polynomial. Then iterative algorithms such as the Newton-Raphson method can be used to estimate θ.

4.3 Simulation study

A simulation study is carried out to evaluate the performance of the quadratic exact likelihood approach (QEL) and the quadratic approximate likelihood approach (QAL) and compare them to the quadratic naive analysis (QNA) which is based on ordinary least squares ignoring measurement error. The quadratic approaches are also compared to their linear counterpart, namely, the linear exact likelihood approach (LEL), the linear approximate likelihood approach (LAL) and the linear naive analysis. Simulations are implemented in the R programming language (R Core Team, 2021).

Data simulation follows a two-step procedure. First, η_i and ξ_i are generated from the quadratic model (4.2) for $\eta_i|\xi_i$ combined with a normal control risk distribution for ξ_i. Binary treatment and control responses are then generated by simulating from model (2.4). The observed risk measures $\hat{\eta}_i$ and $\hat{\xi}_i$ with associated variances $s^2_{\eta_i}$ and $s^2_{\xi_i}$ are then computed based on formulas (2.7) and (2.8), using a continuity correction if needed. We consider a Binomial likelihood rather than a normal likelihood since we want to distinguish between the exact measurement error model and the approximate measurement error model.

Simulation parameters include the the number of studies in the meta-analysis $n \in \{10, 20\}$, the between-study variance $\tau^2 \in \{0, 0.5, 1\}$, and the regression coefficient of the quadratic term $\beta_2 \in \{0, 0.8\}$. A zero of β_2 allows us to evaluate the performance of the quadratic approaches when the true relationship is linear. Other parameters $\left(\beta_0, \beta_1, \mu_\xi, \sigma^2_\xi\right)^\top$ are set to $(0, 1, 0, 1)^\top$. Under every scenario, the number of patients in the treatment group n_{iT} and the number of patients in the control group n_{iC} are generated from a uniform distribution $U(15, 200)$). Each scenario is simulated 1,000 times and the performance of the approaches is evaluated using several metrics: bias, average standard error (se), standard deviation (sd), frequency that the estimator converges (conv), and empirical coverage probability (ecp) (Morris *et al.*, 2019).

Likelihood functions are maximized with the Nelder-Mead algorithm (Nelder and Mead, 1965). Using Gaussian-Hermite quadrature to approximate likelihood functions, the number of nodes in each dimension is set to 20. Estimates from the naive analysis is used as an initial guess for applying the the likelihood approach.

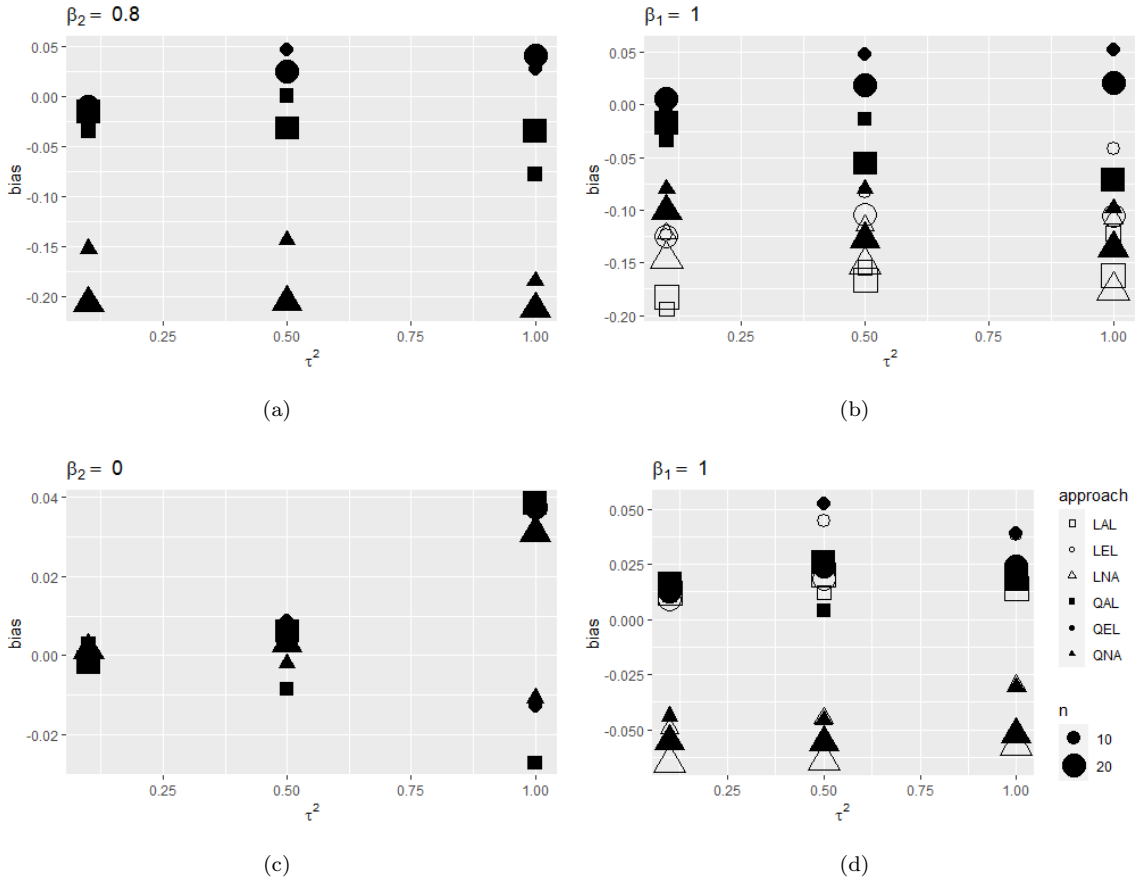

FIGURE 4.1: Bias from the quadratic naive analysis (QNA), the quadratic approximate likelihood (QAL), the quadratic exact likelihood (QEL), the linear naive analysis (LNA), the linear approximate likelihood (LAL), and the linear exact likelihood (LEL). Panels a)-c): interest on β_2, panel b): interest on β_1 with $\beta_2 = 0.8$; panel d): interest on β_1 when $\beta_2 = 0$.

4.3.1 Results

Figure 4.1 shows the bias of the estimators of β_2 and β_1 from QNA, QAL, QEL, LNA, LAL and LEL. There is almost no difference between bias of estimators from QAL and QEL. They outperform the other approaches since their estimators have very small absolute biases, while the linear likelihood approaches underestimate β_1 when the true relationship is quadratic. When the true relationship is linear, the biases of the estimators of β_1 from the quadratic likelihood approaches are approximately equal to the ones from LAL and LEL. Moreover, not accounting for measurement errors in both linear model and quadratic model also results in negative bias.

Figures 4.2–4.3 show the standard errors and the standard deviations of the estimators of β_2 and β_1 from QNA, QAL, QEL, LNA, LAL and LEL. The quadratic likelihood estimators have the largest standard errors because their associated models are the

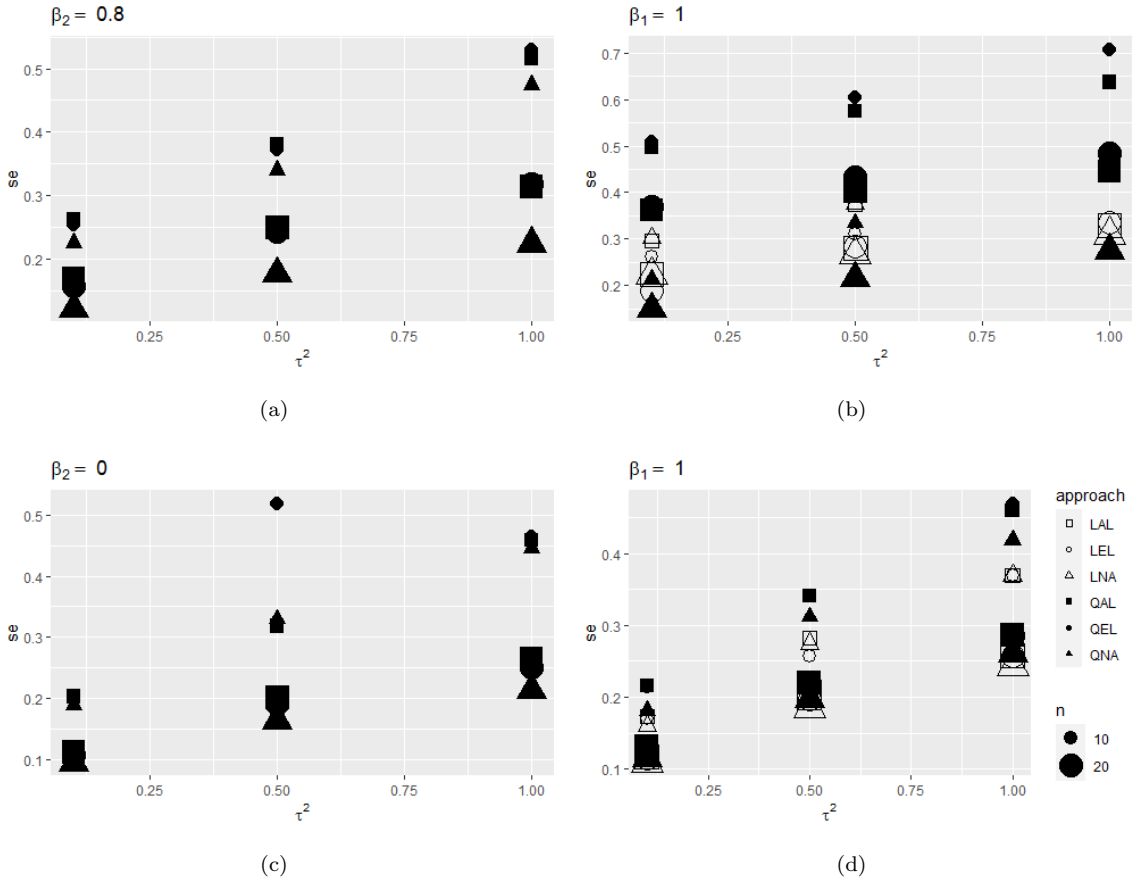

FIGURE 4.2: Standard errors (se) from the quadratic naive analysis (QNA), the quadratic approximate likelihood (QAL), the quadratic exact likelihood (QEL), the linear naive analysis (LNA), the linear approximate likelihood (LAL), and the linear exact likelihood (LEL). Panels a)-c): interest on β_2, panel b): interest on β_1 with $\beta_2 = 0.8$; panel d): interest on β_1 when $\beta_2 = 0$.

most complicated. Furthermore, the standard errors increase when the residual variance grows or the number of studies falls. When the residual variance is small, the standard errors of quadratic likelihood estimators of β_2 are much smaller than the associated standard deviations. This is also true when the number of studies is small and the residual variance is up to moderate. The standard deviations decrease when the number of studies or the residual variance increases.

Panels (a)-(c) report the results for the empirical coverage probability of β_2 when the quadratic models are fitted under $\beta_2 = 0.8$ (Panel (a)) and under $\beta_2 = 0$ (Panel (c)). Results show a satisfactory behavior of the QAL and QEL solutions with respect to alternatives, with values closer to the target 95% level, especially for large τ^2. As expected, correction solutions deeply outperform the quadratic naive analysis (QNA) which does not improve empirical coverage probability when the sample size increases, as a consequence of large bias of the estimators. When focusing on β_1, the improvements

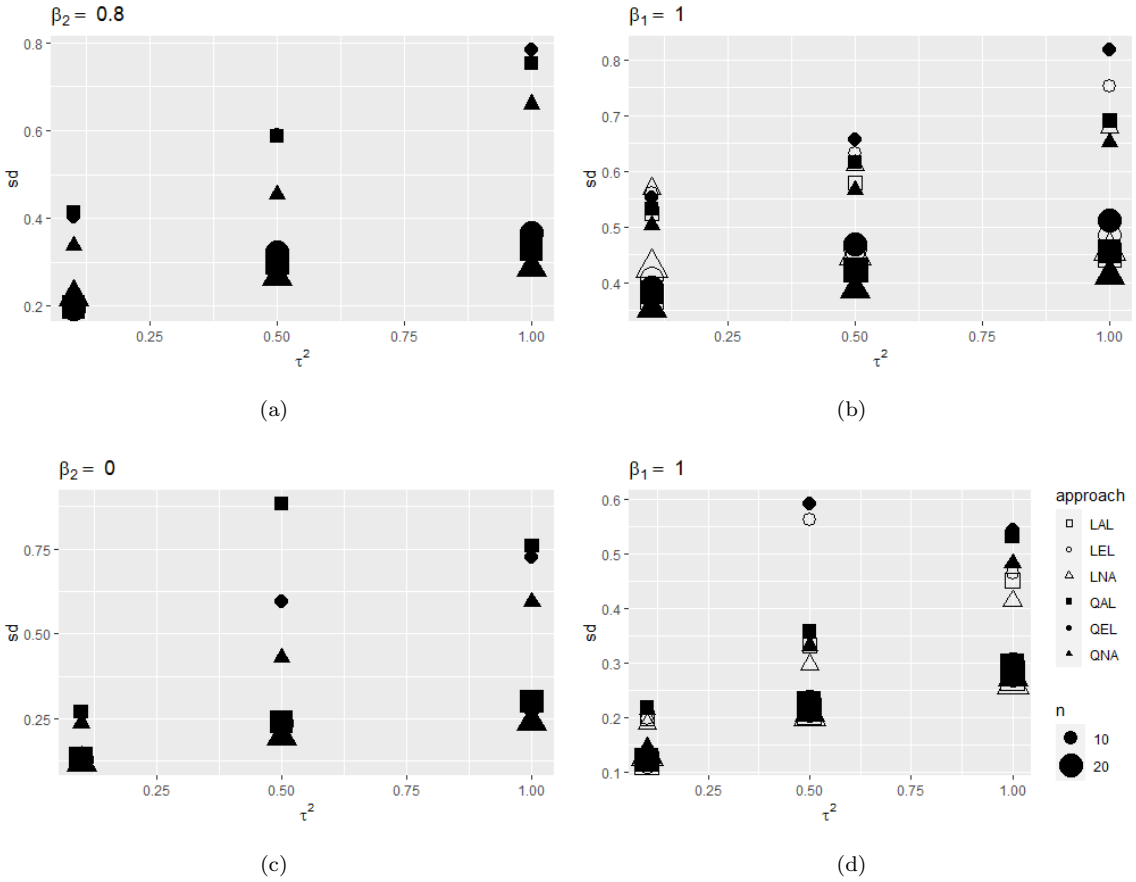

FIGURE 4.3: Standard deviations (sd) from the quadratic naive analysis (QNA), the quadratic approximate likelihood (QAL), the quadratic exact likelihood (QEL), the linear naive analysis (LNA), the linear approximate likelihood (LAL), and the linear exact likelihood (LEL). Panels a)-c): interest on β_2, panel b): interest on β_1 with $\beta_2 = 0.8$; panel d): interest on β_1 when $\beta_2 = 0$.

provided by the U-shaped model with respect to the linear counterpart are relevant, especially when $\beta_2 = 0.8$ (Panel (b)). Increasing the sample size is an expected instrument to make empirical coverage probability closer to the target level.

4.4 Examples

4.4.1 Parkinson's disease dataset

Craft and Watson (2004) and Santiago and Potashkin (2013) have suggested a potential biological association between diabetes mellitus and Parkinson's disease because of their similar pathogenic pathways. Although many systematic reviews and meta-analyses (e.g., Cereda *et al.*, 2011, 2013) have investigated whether having diabetes

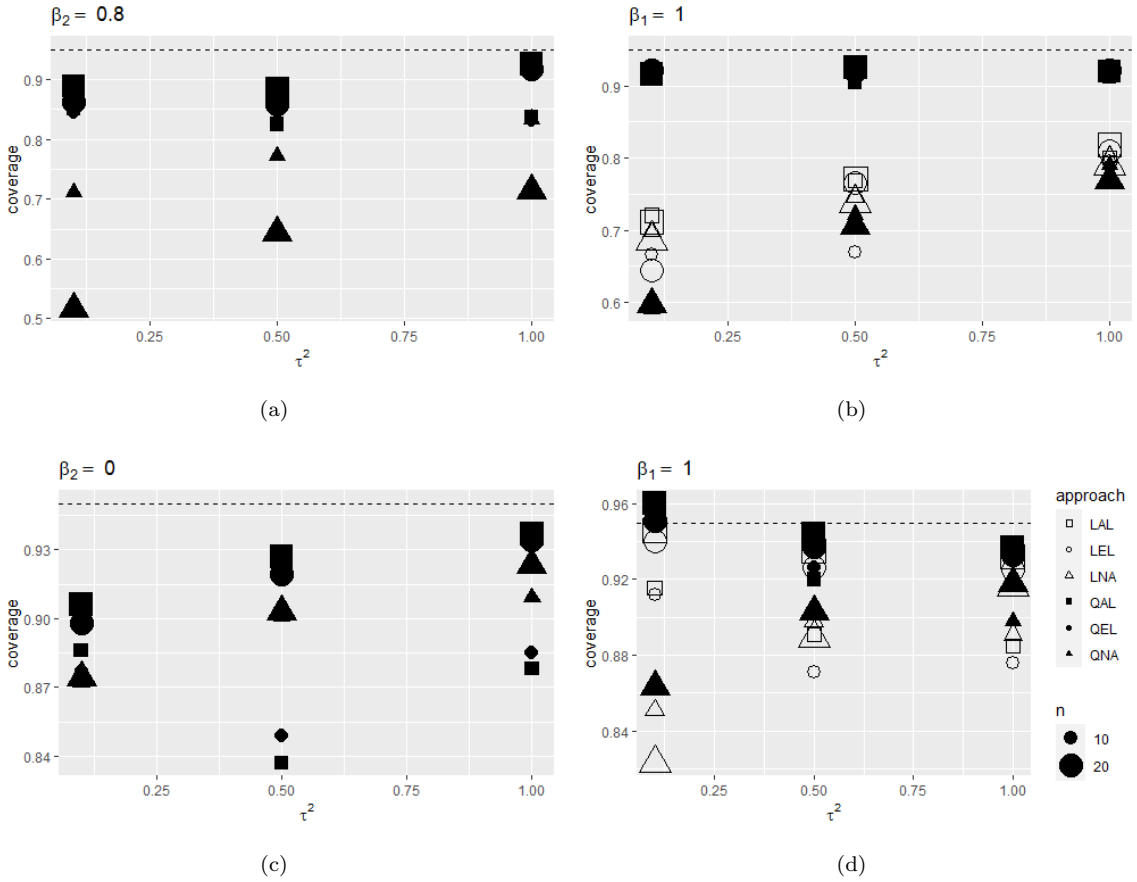

FIGURE 4.4: Empirical coverage probability (ecp) from the quadratic naive analysis (QNA), the quadratic approximate likelihood (QAL), the quadratic exact likelihood (QEL), the linear naive analysis (LNA), the linear approximate likelihood (LAL), and the linear exact likelihood (LEL). Panels a)-c): interest on β_2, panel b): interest on β_1 with $\beta_2 = 0.8$; panel d): interest on β_1 when $\beta_2 = 0$.

increases the risk of future Parkinson's disease, the purported relationship was still unproven. Most recently, Lu *et al.* (2014) conducted a meta-analysis of 14 case-control studies. Table 4.1 reports the number of events, intended as diagnosis of Parkinson's disease, for subjects affected or not affected by type-2 diabetes. Lu *et al.* (2014) analyzed the data using a classical meta-analysis model based on the log odds ratios, resulting in diabetes less likely to occur in patients with Parkinson's disease (overall OR = 0.75; 95% confidence interval = $[0.58, 0.98]$). However, they highlighted substantial heterogeneity among the studies. The association was different in different subgroups of patients stratified by gender, geographic location, source of control group, smoking, antidiabetes drug prescription and duration of diabetes mellitus. This means that control risk regression can be an appropriate instrument for data analysis. Guolo (2022) confirmed the association by fitting a linear control risk regression model using likelihood and non-likelihood based solutions. The likelihood results are shown in the fourth row

of Table 4.2.

TABLE 4.1: Parkinson's disease dataset (Lu *et al.*, 2014).

	Diabetes		Non diabetes	
Study	Events	Total	Events	Total
1	6	35	12	105
2	6	178	58	534
3	18	212	8	175
4	12	74	18	148
5	11	93	26	93
6	13	196	17	196
7	10	249	39	368
8	13	318	31	318
9	18	197	24	197
10	17	228	29	228
11	26	352	61	484
12	48	13695	223	68445
13	126	1931	482	9651
14	291	3637	308	3637

Here, the analysis in Guolo (2022) is extended by fitting quadratic models when the risk measures in the case group and the control group are the log odds of getting Parkinson's disease in the case group and the control group, respectively. Results are reported in Table 4.2 and they include the estimates of $(\beta_0, \beta_1, \beta_2, \tau^2)^\top$ obtained from different linear and quadratic models (QAL, QEL, QNA, LAL, LEL, LNA), and the associated standard errors. There is one study whose risk measures are very different from the ones of the others. Using Gaussian-Hermite quadrature, we set the number of nodes to 20.

All the linear approaches provide no statistically significant association between diabetes and Parkinson's disease as the estimate of β_1 is not different from one, in this way confirming previous results in Lu *et al.* (2014) and Guolo (2022). The estimates of β_1 from the linear-based approaches are much different from those from the quadratic-based counterparts. All the estimates of the intercept are very negative because of centering the risk measure in the control condition. The intercept from all the approaches is also significantly different from zero. Quadratic control risk models with specification of measurement error result in smaller estimates of τ^2 and associated standard errors compared to the ones from linear counterparts, in this way indicating that additional

TABLE 4.2: Estimated parameters and the associated standard errors of linear and quadratic naive analyses, linear and quadratic control risk regression models on Parkinson's disease dataset (Lu *et al.*, 2014).

Approach	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}^2$
QAL	-2.540 (0.162)	0.434 (0.196)	**-0.149 (0.063)**	0.053 (0.072)
QEL	-2.583 (0.164)	0.421 (0.213)	**-0.151 (0.069)**	0.073 (0.070)
QNA	-2.570 (0.182)	0.455 (0.255)	-0.137 (0.094)	0.251 (0.084)
LAL	-2.695 (0.252)	0.805 (0.112)	-	0.133 (0.089)
LEL	-2.730 (0.251)	0.795 (0.116)	-	0.144 (0.092)
LNA	-2.717 (0.158)	0.759 (0.153)	-	0.299 (0.105)

between-study heterogeneity has been taken into account by the U-shaped relationship. Negative estimates of β_2 from QAL and QEL show that the risk of Parkinson's disease in the diabetic group increases when the risk in the group without diabetes grows to a certain level and then decreases after reaching this level. Not surprisingly, the quadratic control risk regression model ignoring the presence of measurement error results in much larger estimates of τ^2 and associated standard errors compared to the alternatives. Moreover, QNA gives no significant association between diabetes and Parkinson's disease, see Table 4.2.

Figure 4.5 reports the graphs of models fitted with QAL, QEL, QNA, LAL, LEL and LNA. The graph shows similar behaviors for the quadratic proposal and a slight difference for the linear proposals as the control risk measure decreases.

4.4.2 Myocardial injury dataset

Re-consider the example of myocardial injury dataset (Sanz-Sánchez *et al.*, 2021) analyzed in Chapter 3 where the inclusion of additional covariates in the classical control risk regression has been considered as a way to better explain between-study heterogeneity. In the previous analysis, associations were found between the treatment risk, the control risk and the log odds of hypertension. Data can be re-analyzed with the quadratic model to understand whether a U-shaped relationship between the treatment risk and the control risk can properly explain heterogeneity across studies. The quadratic control risk regression model is fitted to the dataset using QAL, QEL, and QNA approaches. Furthermore, the linear model is re-fitted to the dataset using LEL and LNA approaches. The number of nodes in Gaussian-Hermite quadrature is set to 20.

The estimates of regression coefficients, the residual variance and their associated standard errors are presented in Table 4.3. Except for the linear approximate likelihood

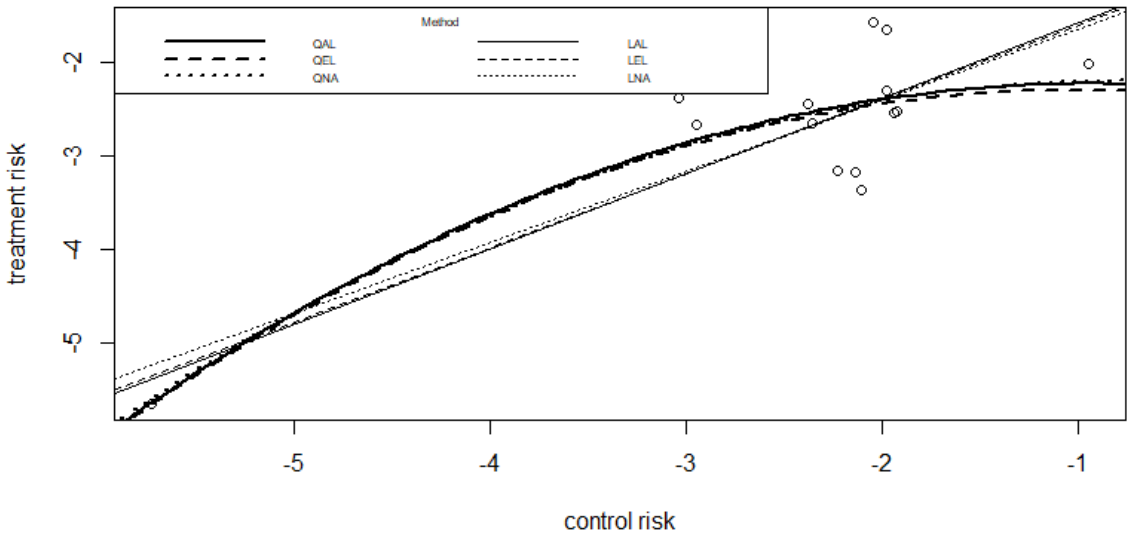

FIGURE 4.5: Parkinson's disease dataset (Lu *et al.*, 2014). LAL (solid straight line), LEL (dashed straight line), QAL (solid straight curve), QEL (dashed straight curve), LNA (dotted line), and QNA (dotted curve).

approach, relationships between the risk measures suggested by other approaches are statistically significant. Although the estimates of β_2 and β_1 from the quadratic approaches are slightly similar, the estimates of the residual variance are much smaller when accounting for measurement error. These estimates are among the smallest and only larger than the one from the linear control risk regression model with the log odds of hypertension. Figure 4.6 shows the graphs of fitted control risk regression models without additional covariates. As suggested by the negative estimates of β_2 and the associated graphs from the quadratic approaches, the treatment risk increases with the control risk until the baseline risk is around -2 and then decreases. Table 4.3 also includes the AICs (Akaike information criterions) of the extended model and the model fitted in Chapter 3 with the log odds of hypertension as a significant additional covariate. From the table, the model providing the smallest AIC is the quadratic model QAL.

TABLE 4.3: Myocardial injury dataset (Sanz-Sánchez *et al.*, 2021). Estimates of $(\beta_0, \beta_1, \beta_2, \tau^2)$, associated standard errors and AICs using QAL, QEL, QNA, LAL, LEL and LNA approaches. Comparison to model with hypertension fitted with the pseudo-likelihood approach. Significant coefficients are highlighted.

Approach	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\tau}^2$	AIC
QAL	0.283 (0.194)	0.236 (0.199)	**-0.231 (0.111)**	0.147 (0.087)	**32.53**
QEL	0.314 (0.204)	0.272 (0.219)	**-0.255 (0.086)**	0.143 (0.083)	116.642
QNA	0.285 (0.287)	0.284 (0.152)	**-0.212 (0.101)**	0.395 (0.132)	-
LAL	0.045 (0.166)	0.229 (0.153)	-	0.203 (0.110)	46.948
LEL	-0.041 (0.201)	**0.329 (0.137)**	-	0.256 (0.132)	117.454
LNA	-0.168 (0.215)	**0.451 (0.147)**	-	0.555 (0.194)	-
LAL+Hypertension	-0.016 (0.180)	**0.329 (0.110)**	**-0.354 (0.150)**	**0.038 (0.049)**	104.053

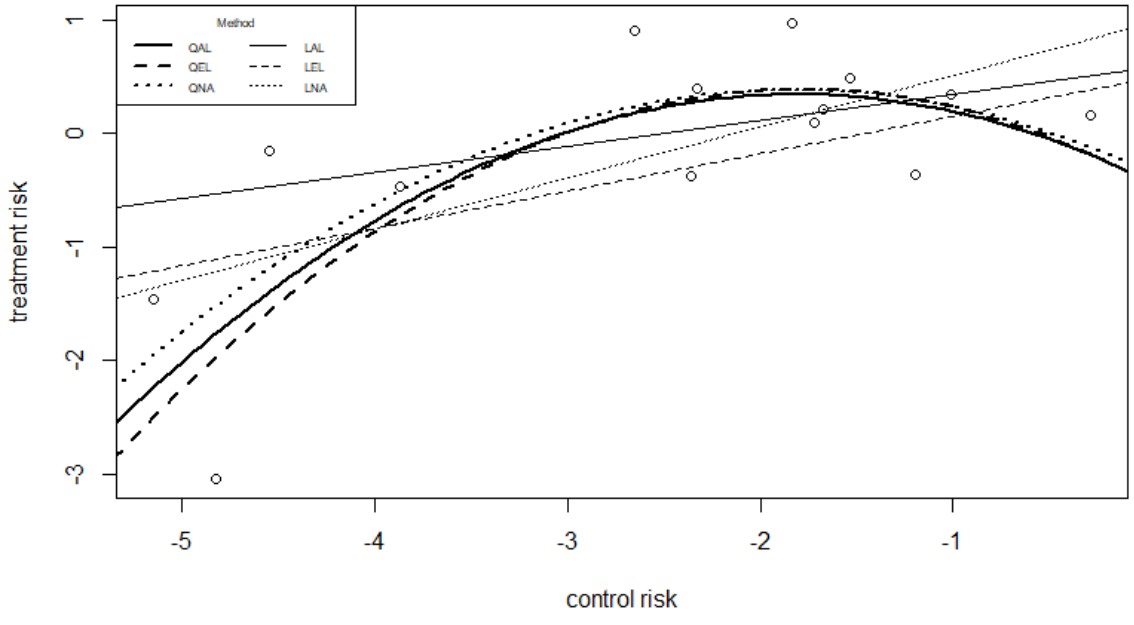

FIGURE 4.6: Myocardial injury dataset (Sanz-Sánchez *et al.*, 2021). LAL (solid straight line), LEL (dashed straight line), QAL (solid straight curve), QEL (dashed straight curve), LNA (dotted line), and QNA (dotted curve).

Conclusions

Discussion

This thesis considered extensions of the control risk regression model used in meta-analysis to evaluate the effectiveness of a treatment in clinical trials comparing a treatment group and a control group. The suggested extensions refer to the inclusion of additional covariates in the classical model which only accounts for the control rate in order to explain between-study heterogeneity.

When the additional study-specific covariates are affected by measurement error, likelihood-based inference is carried out by adopting appropriate measurement error correction solutions, which can result in several issues. The available information at the study level does not provide within-study covariances between risk measures and the covariates. The thesis illustrates how to recover such information exploiting subgroup study-specific summary data. When they cannot be derived, the pseudo-likelihood approach suggested in the thesis is a viable solution. The approach developed under a conditional independence assumption between the observed measures of risk and covariates gives rise to a pseudo-likelihood function with a closed-form expression under the approximate measurement error model. The proposed solutions perform satisfactorily in a series of simulations under different conditions. The pseudo-likelihood approach has the advantage of being less computationally demanding than the likelihood approach based on subgroup summary information, when available, while maintaining small bias of the estimators and empirical coverage probabilities close to the target level.

A second extension of the classical model considered the inclusion of a quadratic term associated to the control rate as an additional covariate, in order to make the relationship between the treatment risk and the control risk more flexible, able to handle more complex, as U-shaped, situations. The likelihood approach, under both the approximate within-study distribution and the exact within-study distribution, gives rise to a likelihood function without closed-form expression, which can be faced through numerical integration.

The performance of the quadratic likelihood-based solutions has been investigated in a

series of simulations, showing a satisfactory performance in terms of accuracy of inferential results and with convergence problems rarely experienced.
The extensions of the classical control risk regression model examined in the thesis have been applied to real meta-analyses of medical interest. Results are encouraging, suggesting that accounting for additional study-specific information and/or more flexible models can help in explaining the unobserved between-study heterogeneity. The R code useful to run the analyses according to the approaches in Chapter 3 and in Chapter 4 is reported in the Appendix, Section B.

Future directions of research

The studies carried out in the thesis leave space for future different lines of research. The considered likelihood-based approach to measurement error correction is a structural solution, that requires the specification of a model for the unobserved risk measures. Functional models developed in the measurement error literature can be considered as an interesting alternative, which do not require any assumptions for the underlying quantities. For example, an interesting extension of the work might consider the use of corrected scores and conditional scores, that have been previously applied in the classical control risk regression in Ghidey *et al.* (2013). In addition, the use of SIMEX (Carroll *et al.*, 2006), a simulation-based approach for measurement error correction, can be a promising alternative, given previous results in Guolo (2014) in the classical control risk regression, although the computation cost might be relevant when the number of study-specific covariates increases.
When examining the quadratic relationship between the treatment risk and the control risk, the present work has been carried out from a frequentist point of view. An interesting future research may develop analysis from a Bayesian perspective, under an uninformative prior on the term associated to the quadratic component. At the time of writing, in fact, the literature in control risk regression considers Bayesian analysis only in case of linear relationship between the risk measures (see, e.g., Arends *et al.*, 2000, Lee and Thompson, 2008). In addition, future developments can refer to two drawbacks associated to the quadratic regression models, namely, poor extrapolation at extreme values of predictors and inability to fit data with several thresholds, as noted in Bagnardi *et al.* (2004). More complex and flexible relationships between the treatment risk and the control risk can be considered rather than the quadratic one, in order to better describe their relationship, as, for example, polynomial with fractional degree or spline solutions. The choice, however, should be concerned with the small number of

studies, which is a typical feature of meta-analysis, and that can be associated to the risk of overfitting.

Appendix

Further simulation results of chapter 3

TABLE A.1: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2\right)^\top$ and number of convergent solutions over 1,000 replicates for the uncorrected approach and the likelihood approach when data follows scenario 1. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

τ^2	par.	appr.	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.1	β_0	lik	0.030	0.255	0.317	988	0.037	0.188	0.201	994
		naive	-0.085	0.244	0.283	1000	-0.134	0.172	0.189	1000
	β_1	lik	0.059	0.321	0.392	988	0.049	0.236	0.248	994
		naive	-0.107	0.259	0.319	1000	-0.173	0.170	0.214	1000
	β_2	lik	-0.002	0.137	0.157	988	-0.012	0.080	0.083	994
		naive	0.013	0.147	0.176	1000	0.017	0.095	0.096	1000
	μ_ξ	lik	0.006	0.189	0.211	988	0.025	0.138	0.150	994
	τ^2	lik	-0.055	0.051	0.061	981	-0.041	0.051	0.056	987
		naive	0.072	0.065	0.109	1000	0.149	0.073	0.106	1000
	σ_ξ^2	lik	-0.057	0.177	0.196	988	-0.067	0.129	0.142	994
0.5	β_0	lik	0.083	0.460	0.627	997	0.053	0.319	0.363	1000
		naive	-0.077	0.431	0.579	1000	-0.135	0.303	0.331	1000
	β_1	lik	0.113	0.586	0.796	997	0.084	0.386	0.454	1000
		naive	-0.084	0.449	0.645	1000	-0.162	0.286	0.350	1000
	β_2	lik	-0.022	0.172	0.217	997	-0.021	0.202	0.221	1000
		naive	0.021	0.233	0.257	1000	0.021	0.300	0.292	1000
	μ_ξ	lik	0.010	0.188	0.202	997	0.031	0.138	0.145	1000
	τ^2	lik	-0.240	0.185	0.218	997	-0.124	0.171	0.183	999
		naive	0.055	0.208	0.323	1000	0.142	0.187	0.254	1000
	σ_ξ^2	lik	-0.066	0.174	0.192	997	-0.072	0.129	0.139	1000
1	β_0	lik	0.138	0.650	0.894	998	0.064	0.426	0.488	1000
		naive	-0.043	0.579	0.767	1000	-0.142	0.392	0.455	1000
	β_1	lik	0.182	0.831	1.212	998	0.109	0.522	0.681	1000
		naive	-0.047	0.626	0.872	1000	-0.165	0.388	0.473	1000
	β_2	lik	-0.024	0.245	0.306	998	-0.023	0.237	0.265	1000
		naive	0.035	0.305	0.421	1000	0.029	0.289	0.335	1000
	μ_ξ	lik	0.026	0.184	0.207	998	0.031	0.138	0.142	1000
	τ^2	lik	-0.446	0.346	0.417	995	-0.261	0.304	0.326	1000
		naive	0.085	0.406	0.670	1000	0.118	0.326	0.400	1000
	σ_ξ^2	lik	-0.081	0.167	0.186	998	-0.074	0.128	0.134	1000

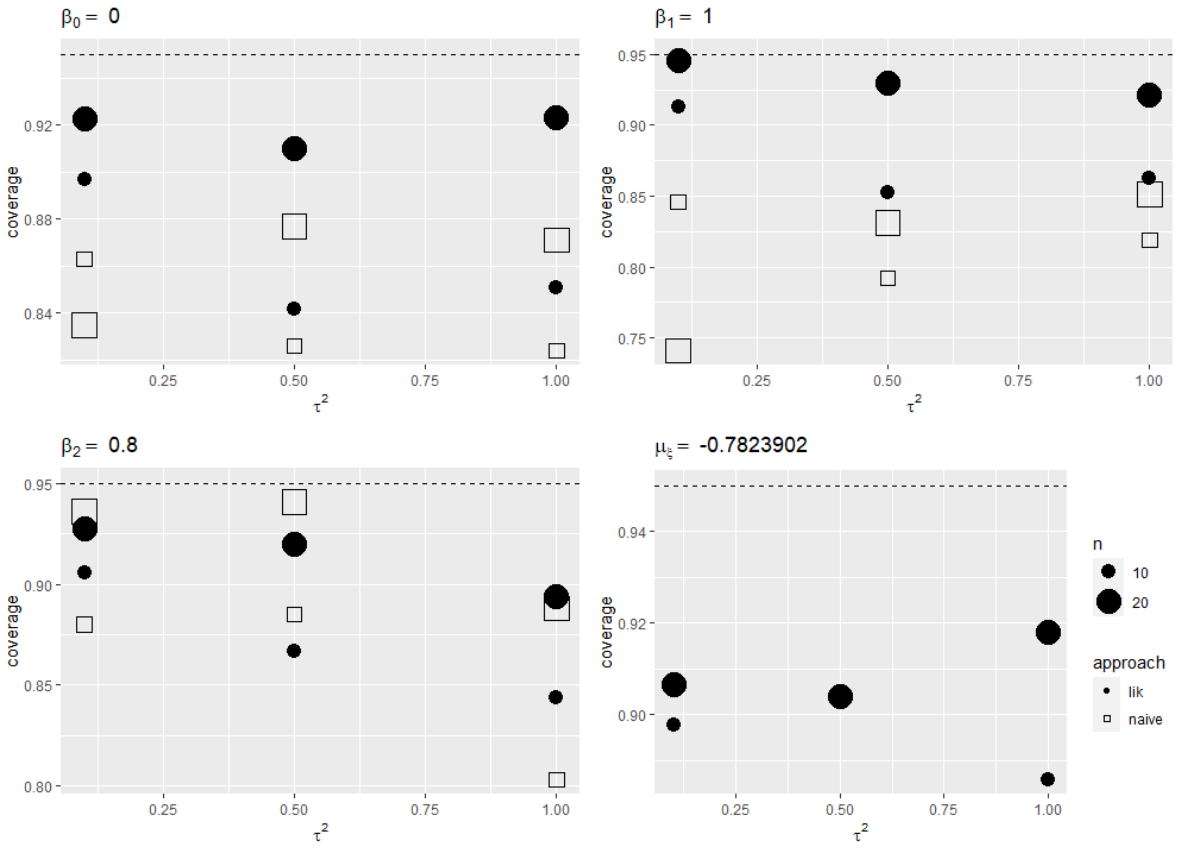

FIGURE A.1: Empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi)^\top$ for the uncorrected approach and the likelihood approach when data follows scenario 1. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

TABLE A.2: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2\right)^\top$ and number of convergent solutions over 1,000 replicates for the uncorrected approach and the likelihood approach when data follows scenario 2. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

τ^2	par.	appr.	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.1	β_0	lik	0.028	0.271	0.341	986	0.045	0.209	0.219	996
		naive	-0.097	0.255	0.309	1000	-0.154	0.208	0.219	1000
	β_1	lik	0.029	0.335	0.461	986	0.068	0.227	0.243	996
		naive	-0.134	0.261	0.326	1000	-0.199	0.166	0.203	1000
	β_2	lik	-0.034	0.277	0.336	986	-0.013	0.194	0.212	996
		naive	-0.012	0.319	0.365	1000	0.010	0.217	0.243	1000
	μ_ξ	lik	0.016	0.184	0.205	986	0.019	0.138	0.142	996
	τ^2	lik	-0.050	0.052	0.063	976	-0.034	0.052	0.059	993
		naive	0.077	0.066	0.111	1000	0.127	0.066	0.095	1000
	σ_ξ^2	lik	-0.076	0.167	0.190	986	-0.063	0.129	0.138	996
0.5	β_0	lik	0.086	0.480	0.648	993	0.100	0.408	0.456	1000
		naive	-0.101	0.478	0.541	1000	-0.134	0.418	0.491	1000
	β_1	lik	0.142	0.564	0.822	993	0.129	0.389	0.448	1000
		naive	-0.104	0.466	0.558	1000	-0.177	0.284	0.345	1000
	β_2	lik	-0.006	0.430	0.516	996	-0.028	0.361	0.392	1000
		naive	0.035	0.509	0.569	1000	0.013	0.417	0.487	1000
	μ_ξ	lik	0.025	0.182	0.200	996	0.031	0.137	0.145	1000
	τ^2	lik	-0.216	0.188	0.221	993	-0.120	0.174	0.192	1000
		naive	0.007	0.190	0.303	1000	0.121	0.181	0.233	1000
	σ_ξ^2	lik	-0.086	0.164	0.184	994	-0.075	0.127	0.134	1000
1	β_0	lik	0.052	0.592	0.767	997	0.088	0.503	0.573	1000
		naive	-0.107	0.587	0.741	1000	-0.153	0.516	0.549	1000
	β_1	lik	0.081	0.732	1.008	997	0.100	0.505	0.590	1000
		naive	-0.127	0.635	0.807	1000	-0.189	0.375	0.455	1000
	β_2	lik	-0.035	0.932	1.218	999	-0.057	0.445	0.461	1000
		naive	0.032	0.865	1.324	1000	0.014	0.518	0.536	1000
	μ_ξ	lik	0.018	0.186	0.202	999	0.024	0.138	0.146	1000
	τ^2	lik	-0.405	0.350	0.419	997	-0.248	0.305	0.321	1000
		naive	-0.061	0.351	0.546	1000	0.074	0.313	0.402	1000
	σ_ξ^2	lik	-0.071	0.173	0.196	999	-0.073	0.128	0.143	1000

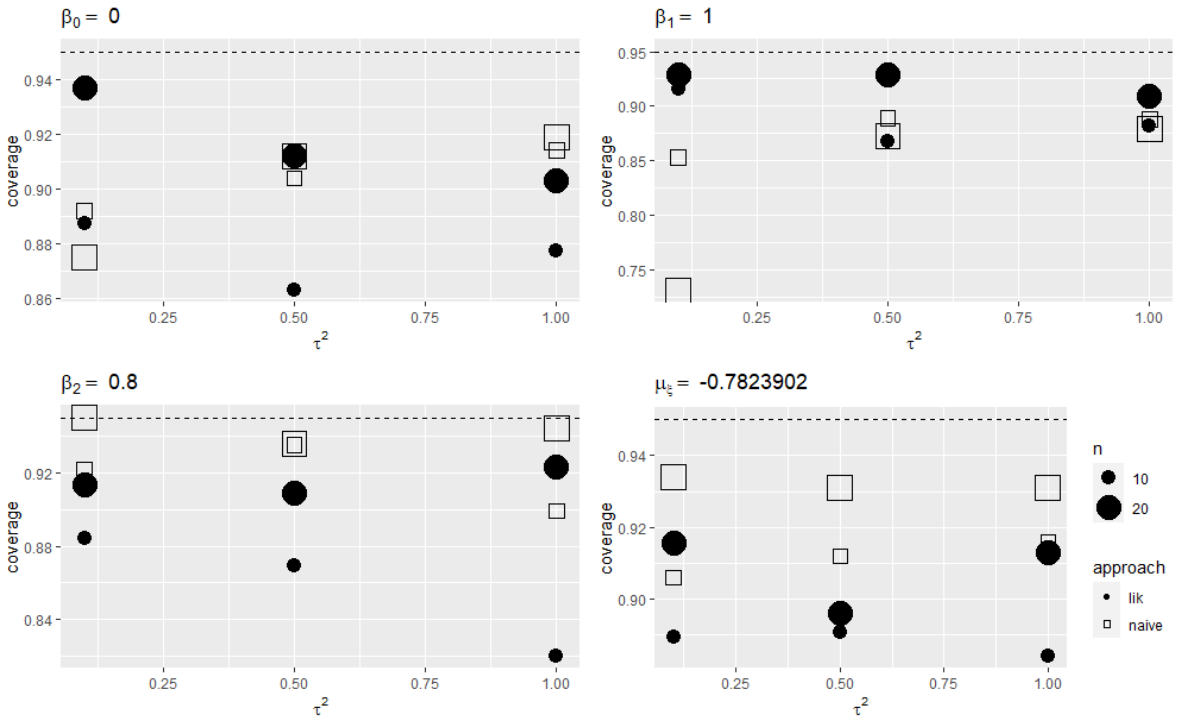

FIGURE A.2: Empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi)^\top$ for the uncorrected approach and the likelihood approach when data follows scenario 2. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

TABLE A.3: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 4 and $\tau^2 = 0.5$. Underlying risk distributed as a standard normal.

τ^2	par.	approach	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.5	β_0	lik	0.006	0.233	0.282	998	0.013	0.172	0.193	1000
		pseudo-lik	0.006	0.232	0.282	1000	0.013	0.172	0.194	1000
		naive	0.001	0.265	0.391	1000	0.006	0.190	0.226	1000
	β_1	lik	0.021	0.288	0.341	998	0.019	0.204	0.227	1000
		pseudo-lik	0.020	0.286	0.342	1000	0.018	0.204	0.226	1000
		naive	-0.029	0.291	0.407	1000	-0.075	0.186	0.223	1000
	β_2	lik	-0.013	0.258	0.321	998	-0.019	0.181	0.204	1000
		pseudo-lik	-0.013	0.257	0.320	1000	-0.019	0.182	0.202	1000
		naive	0.012	0.306	0.427	1000	0.010	0.198	0.234	1000
	μ_ξ	lik	0.005	0.298	0.318	998	-0.008	0.220	0.231	1000
		pseudo-lik	0.004	0.299	0.316	1000	-0.008	0.220	0.231	1000
	μ_ζ	lik	0.010	0.295	0.325	998	-0.006	0.217	0.221	1000
		pseudo-lik	0.008	0.295	0.324	1000	-0.006	0.216	0.220	1000
	τ^2	lik	-0.203	0.195	0.235	996	-0.129	0.169	0.180	999
		pseudo-lik	-0.204	0.194	0.237	1000	-0.128	0.170	0.180	1000
		naive	0.126	0.234	0.397	1000	0.130	0.184	0.244	1000
	σ_ξ^2	lik	-0.115	0.432	0.425	998	-0.090	0.324	0.314	1000
		pseudo-lik	-0.113	0.434	0.428	1000	-0.089	0.325	0.315	1000
	σ_ζ^2	lik	-0.082	0.415	0.446	998	-0.043	0.305	0.309	1000
		pseudo-lik	-0.080	0.416	0.447	1000	-0.045	0.305	0.306	1000

TABLE A.4: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 4 and $\tau^2 = 1$. Underlying risk distributed as a standard normal.

τ^2	par.	approach	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
1	β_0	lik	-0.014	0.312	0.397	1000	0.003	0.226	0.240	1000
		pseudo-lik	-0.015	0.313	0.393	999	0.004	0.226	0.241	1000
		naive	-0.030	0.370	0.465	1000	0.002	0.250	0.327	1000
	β_1	lik	-0.008	0.378	0.488	1000	-0.017	0.266	0.293	1000
		pseudo-lik	-0.009	0.378	0.485	999	-0.017	0.266	0.293	1000
		naive	-0.057	0.397	0.511	1000	-0.116	0.246	0.325	1000
	β_2	lik	-0.008	0.340	0.450	1000	-0.027	0.241	0.256	1000
		pseudo-lik	-0.009	0.340	0.446	999	-0.026	0.241	0.255	1000
		naive	0.041	0.408	0.519	1000	0.007	0.269	0.352	1000
	μ_ξ	lik	-0.006	0.298	0.325	1000	0.006	0.218	0.226	1000
		pseudo-lik	-0.006	0.298	0.324	999	0.006	0.218	0.226	1000
	μ_ζ	lik	-0.013	0.293	0.316	1000	0.012	0.214	0.220	1000
		pseudo-lik	-0.012	0.293	0.317	999	0.013	0.214	0.220	1000
	τ^2	lik	-0.382	0.355	0.424	1000	-0.252	0.299	0.321	1000
		pseudo-lik	-0.383	0.355	0.422	999	-0.252	0.299	0.322	1000
		naive	-0.029	0.363	0.563	1000	0.159	0.338	0.448	1000
	σ_ξ^2	lik	-0.111	0.436	0.478	1000	-0.108	0.320	0.321	1000
		pseudo-lik	-0.112	0.435	0.475	999	-0.108	0.320	0.318	1000
	σ_ζ^2	lik	-0.096	0.408	0.426	1000	-0.067	0.297	0.314	1000
		pseudo-lik	-0.098	0.407	0.424	999	-0.066	0.298	0.316	1000

TABLE A.5: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 4 and $\tau^2 = 0.1$. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

τ^2	par.	cov	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv
0.1	β_0	lik	0.037	0.289	0.325	984	0.031	0.202	0.229	995
		pseudo-lik	0.038	0.288	0.318	991	0.030	0.201	0.228	994
		naive	-0.100	0.256	0.327	1000	-0.127	0.185	0.225	1000
	β_1	lik	0.050	0.355	0.413	985	0.041	0.250	0.272	995
		pseudo-lik	0.051	0.354	0.406	991	0.039	0.250	0.269	993
		naive	-0.130	0.274	0.374	1000	-0.155	0.185	0.247	1000
	β_2	lik	-0.013	0.160	0.192	986	-0.020	0.120	0.131	997
		pseudo-lik	-0.014	0.160	0.192	992	-0.019	0.120	0.132	995
		naive	0.007	0.173	0.232	1000	0.015	0.126	0.156	1000
	μ_ξ	lik	0.023	0.187	0.203	986	0.028	0.137	0.142	997
		pseudo-lik	0.022	0.187	0.203	992	0.028	0.137	0.144	995
	μ_ζ	lik	0.001	0.295	0.308	986	0.003	0.215	0.227	997
		pseudo-lik	0.001	0.295	0.307	992	0.003	0.215	0.227	995
	τ^2	lik	-0.050	0.058	0.060	979	-0.039	0.058	0.058	980
		pseudo-lik	-0.051	0.058	0.060	986	-0.039	0.058	0.057	974
		naive	0.105	0.077	0.134	1000	0.184	0.083	0.117	1000
	σ_ξ^2	lik	-0.068	0.174	0.188	985	-0.069	0.128	0.136	997
		pseudo-lik	-0.069	0.174	0.186	991	-0.070	0.127	0.134	995
	σ_ζ^2	lik	-0.088	0.413	0.428	986	-0.056	0.302	0.308	997
		pseudo-lik	-0.086	0.414	0.431	992	-0.055	0.302	0.308	995

TABLE A.6: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 4 and $\tau^2 = 0.5$. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

τ^2	par.	cov	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.5	β_0	lik	0.086	0.473	0.602	999	0.060	0.320	0.353	999
		pseudo-lik	0.086	0.471	0.605	998	0.058	0.318	0.349	999
		naive	-0.078	0.444	0.561	1000	-0.139	0.294	0.378	1000
	β_1	lik	0.118	0.583	0.746	999	0.073	0.387	0.438	999
		pseudo-lik	0.118	0.580	0.738	998	0.070	0.384	0.427	999
		naive	-0.094	0.468	0.604	1000	-0.165	0.293	0.397	1000
	β_2	lik	-0.013	0.256	0.328	999	-0.026	0.185	0.214	999
		pseudo-lik	-0.013	0.256	0.327	998	-0.026	0.185	0.215	999
		naive	0.016	0.301	0.387	1000	0.027	0.202	0.269	1000
	μ_ξ	lik	0.011	0.185	0.204	999	0.026	0.138	0.139	999
		pseudo-lik	0.011	0.185	0.203	998	0.026	0.138	0.139	999
	μ_ζ	lik	0.026	0.294	0.311	999	-0.012	0.215	0.218	999
		pseudo-lik	0.027	0.294	0.311	998	-0.011	0.215	0.218	1000
	τ^2	lik	-0.217	0.195	0.234	999	-0.152	0.176	0.190	997
		pseudo-lik	-0.218	0.196	0.233	994	-0.153	0.175	0.190	999
		naive	0.041	0.202	0.334	1000	0.202	0.205	0.282	1000
	σ_ξ^2	lik	-0.074	0.171	0.196	999	-0.066	0.130	0.132	999
		pseudo-lik	-0.075	0.171	0.195	998	-0.065	0.130	0.132	999
	σ_ζ^2	lik	-0.092	0.410	0.435	999	-0.054	0.302	0.310	999
		pseudo-lik	-0.090	0.411	0.437	998	-0.054	0.302	0.311	1000

TABLE A.7: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 4 and $\tau^2 = 1$. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

τ^2	par.	cov	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
1	β_0	lik	0.066	0.617	0.786	999	0.080	0.422	0.462	1000
		pseudo-lik	0.075	0.628	0.796	999	0.082	0.423	0.462	1000
		naive	-0.129	0.600	0.732	1000	-0.116	0.395	0.492	1000
	β_1	lik	0.082	0.755	0.987	999	0.085	0.498	0.537	1000
		pseudo-lik	0.095	0.773	1.004	999	0.088	0.499	0.536	1000
		naive	-0.140	0.635	0.790	1000	-0.131	0.394	0.516	1000
	β_2	lik	-0.008	0.345	0.433	1000	-0.051	0.239	0.258	1000
		pseudo-lik	-0.008	0.346	0.432	999	-0.051	0.239	0.259	1000
		naive	0.023	0.416	0.501	1000	0.005	0.266	0.325	1000
	μ_ξ	lik	0.019	0.183	0.203	1000	0.023	0.137	0.144	1000
		pseudo-lik	0.019	0.183	0.203	999	0.023	0.137	0.144	1000
	μ_ζ	lik	0.008	0.289	0.313	1000	0.003	0.217	0.216	1000
		pseudo-lik	0.008	0.289	0.312	999	0.003	0.217	0.217	1000
	τ^2	lik	-0.401	0.361	0.410	997	-0.297	0.306	0.332	999
		pseudo-lik	-0.399	0.363	0.413	997	-0.297	0.307	0.332	999
		naive	-0.029	0.363	0.559	1000	0.198	0.349	0.461	1000
	σ_ξ^2	lik	-0.082	0.168	0.186	999	-0.072	0.128	0.129	1000
		pseudo-lik	-0.083	0.168	0.185	999	-0.071	0.128	0.130	1000
	σ_ζ^2	lik	-0.125	0.395	0.423	1000	-0.038	0.307	0.310	1000
		pseudo-lik	-0.126	0.394	0.421	999	-0.038	0.307	0.310	1000

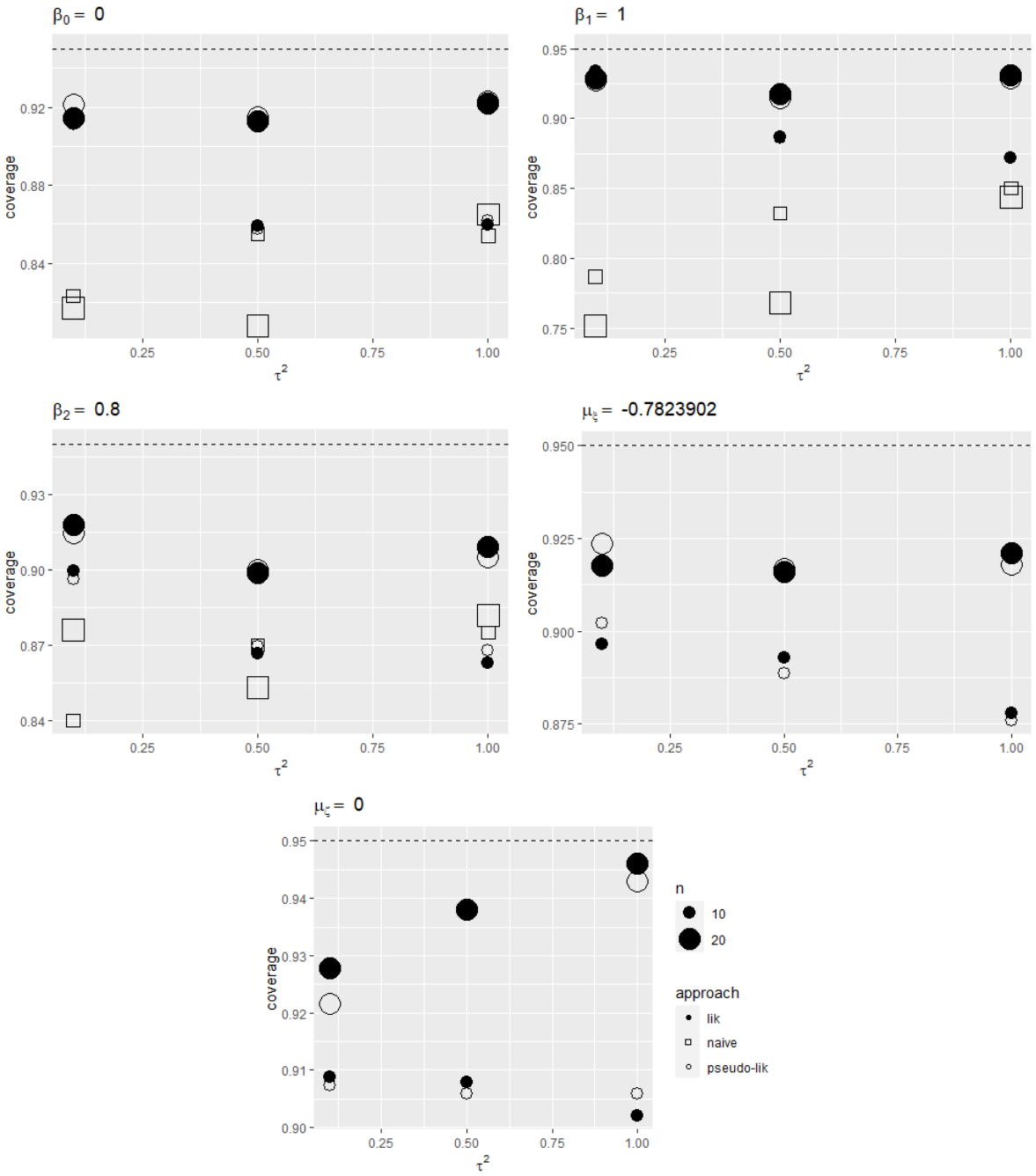

FIGURE A.3: Empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta)^\top$ for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 4. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

TABLE A.8: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 3 and $\tau^2 = 0.5$. Underlying risk distributed as a standard normal.

τ^2	par.	cov	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.5	β_0	lik	0.005	0.236	0.279	999	-0.006	0.169	0.183	1000
		pseudo-likelihood	0.004	0.237	0.281	1000	-0.006	0.170	0.184	1000
		naive	0.003	0.273	0.342	1000	-0.006	0.189	0.206	1000
	β_1	lik	-0.009	0.286	0.347	999	0.001	0.196	0.208	1000
		pseudo-likelihood	-0.009	0.287	0.347	1000	0.002	0.196	0.208	1000
		naive	-0.059	0.291	0.367	1000	-0.071	0.186	0.221	1000
	β_2	lik	-0.003	0.272	0.339	999	0.001	0.184	0.200	1000
		pseudo-likelihood	-0.004	0.273	0.337	1000	0.002	0.184	0.199	1000
		naive	-0.010	0.306	0.388	1000	0.001	0.195	0.220	1000
	μ_ξ	lik	0.013	0.297	0.323	999	0.002	0.218	0.231	1000
		pseudo-likelihood	0.012	0.298	0.324	1000	0.002	0.219	0.232	1000
	μ_ζ	lik	-0.004	0.291	0.305	999	-0.003	0.216	0.211	1000
		pseudo-likelihood	-0.003	0.293	0.308	1000	-0.002	0.217	0.212	1000
	τ^2	lik	-0.197	0.199	0.242	994	-0.127	0.165	0.175	1000
		pseudo-likelihood	-0.197	0.201	0.240	997	-0.128	0.166	0.175	1000
		naive	0.082	0.218	0.355	1000	0.119	0.180	0.233	1000
	σ_ξ^2	lik	-0.122	0.429	0.452	999	-0.089	0.317	0.304	1000
		pseudo-likelihood	-0.113	0.434	0.459	1000	-0.082	0.320	0.305	1000
	σ_ζ^2	lik	-0.128	0.408	0.441	999	-0.070	0.308	0.317	1000
		pseudo-likelihood	-0.114	0.414	0.449	1000	-0.058	0.311	0.320	1000

TABLE A.9: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 3 and $\tau^2 = 1$. Underlying risk distributed as a standard normal.

τ^2	par.	cov	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
1	β_0	lik	-0.002	0.307	0.376	999	-0.002	0.228	0.252	1000
		pseudo-likelihood	-0.003	0.308	0.378	999	-0.002	0.229	0.253	1000
		naive	-0.007	0.367	0.451	1000	-0.004	0.252	0.307	1000
	β_1	lik	0.002	0.368	0.472	999	0.007	0.264	0.299	1000
		pseudo-likelihood	0.003	0.368	0.471	999	0.009	0.264	0.298	1000
		naive	-0.044	0.396	0.514	1000	-0.051	0.253	0.321	1000
	β_2	lik	-0.015	0.343	0.431	999	-0.021	0.246	0.267	1000
		pseudo-likelihood	-0.014	0.343	0.434	999	-0.021	0.246	0.267	1000
		naive	-0.000	0.397	0.495	1000	-0.007	0.260	0.318	1000
	μ_ξ	lik	-0.005	0.295	0.302	999	0.003	0.217	0.229	1000
		pseudo-likelihood	-0.004	0.297	0.303	999	0.002	0.218	0.231	1000
	μ_ζ	lik	-0.002	0.296	0.318	999	-0.001	0.217	0.224	1000
		pseudo-likelihood	-0.002	0.298	0.319	999	-0.002	0.218	0.224	1000
	τ^2	lik	-0.414	0.347	0.398	998	-0.227	0.306	0.330	1000
		pseudo-likelihood	-0.413	0.349	0.397	997	-0.223	0.308	0.334	1000
		naive	-0.029	0.363	0.540	1000	0.142	0.333	0.445	1000
	σ_ξ^2	lik	-0.134	0.424	0.431	999	-0.102	0.314	0.319	1000
		pseudo-likelihood	-0.128	0.427	0.436	999	-0.094	0.316	0.323	1000
	σ_ζ^2	lik	-0.100	0.421	0.433	999	-0.060	0.311	0.309	1000
		pseudo-likelihood	-0.088	0.426	0.440	999	-0.047	0.315	0.313	1000

TABLE A.10: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 3 and $\tau^2 = 0.1$. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

τ^2	par.	cov	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.1	β_0	lik	0.038	3.301	0.341	994	0.040	0.207	0.217	993
		pseudo-lik	0.032	0.275	0.332	991	0.036	0.210	0.218	994
	β_1	lik	0.047	4.815	0.449	994	0.061	0.260	0.278	993
		pseudo-lik	0.039	0.340	0.426	992	0.060	0.261	0.282	994
	β_2	lik	0.001	0.168	0.195	994	-0.010	0.125	0.134	994
		pseudo-lik	0.003	0.172	0.192	993	-0.008	0.127	0.134	996
	μ_ξ	lik	0.022	0.190	0.200	994	0.031	0.138	0.150	995
		pseudo-lik	0.018	0.191	0.201	993	0.025	0.139	0.149	996
	μ_ζ	lik	0.010	0.298	0.325	994	0.017	0.217	0.214	995
		pseudo-lik	0.007	0.300	0.328	993	0.020	0.218	0.218	996
	τ^2	lik	-0.053	0.056	0.060	985	-0.032	0.062	0.061	981
		pseudo-lik	-0.054	0.057	0.062	980	-0.037	0.061	0.060	986
	σ_ξ^2	lik	-0.057	0.179	0.189	993	-0.068	0.131	0.135	994
		pseudo-lik	-0.055	0.180	0.189	993	-0.064	0.132	0.140	995
	σ_ζ^2	lik	-0.085	0.429	0.453	994	-0.069	0.313	0.302	995
		pseudo-lik	-0.071	0.436	0.466	993	-0.060	0.316	0.303	996

TABLE A.11: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 3 and $\tau^2 = 0.5$. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

τ^2	par.	cov	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
0.5	β_0	lik	0.064	0.449	0.564	998	0.076	0.329	0.387	1000
		pseudo-lik	0.059	0.454	0.546	998	0.072	0.332	0.382	999
	β_1	lik	0.101	0.561	0.749	998	0.092	0.402	0.477	1000
		pseudo-lik	0.091	0.564	0.690	998	0.089	0.403	0.472	999
	β_2	lik	-0.007	0.267	0.346	1000	-0.017	0.191	0.211	1000
		pseudo-lik	-0.007	0.269	0.346	999	-0.016	0.192	0.211	1000
	μ_ξ	lik	0.025	0.188	0.198	1000	0.035	0.137	0.145	1000
		pseudo-lik	0.021	0.189	0.199	999	0.030	0.138	0.146	1000
	μ_ζ	lik	-0.005	0.294	0.330	1000	-0.005	0.217	0.219	1000
		pseudo-lik	-0.003	0.296	0.331	999	-0.003	0.218	0.222	1000
	τ^2	lik	-0.213	0.197	0.234	996	-0.156	0.176	0.188	999
		pseudo-lik	-0.213	0.200	0.235	997	-0.154	0.179	0.191	998
	σ_ξ^2	lik	-0.065	0.176	0.192	999	-0.079	0.128	0.132	1000
		pseudo-lik	-0.062	0.177	0.194	998	-0.075	0.129	0.134	999
	σ_ζ^2	lik	-0.115	0.415	0.418	1000	-0.073	0.311	0.313	1000
		pseudo-lik	-0.105	0.420	0.423	999	-0.057	0.316	0.318	1000

TABLE A.12: Bias, standard errors (se), standard deviations (sd) of the maximum likelihood estimators of $\left(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\zeta, \tau^2, \sigma_\xi^2, \sigma_\zeta^2\right)^\top$ and numbers of convergent solutions over 1,000 replicates for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 3 and $\tau^2 = 1$. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

τ^2	par.	cov	$n = 10$				$n = 20$			
			bias	se	sd	conv.	bias	se	sd	conv.
1	β_0	lik	0.051	0.592	0.737	1000	0.115	0.427	0.495	999
		pseudo-lik	0.052	0.586	0.744	998	0.115	0.431	0.519	999
	β_1	lik	0.050	0.728	0.937	1000	0.113	0.512	0.605	999
		pseudo-lik	0.051	0.704	0.937	998	0.114	0.515	0.647	999
	β_2	lik	-0.014	0.346	0.436	1000	-0.029	0.248	0.277	1000
		pseudo-lik	-0.016	0.346	0.433	1000	-0.030	0.248	0.275	1000
	μ_ξ	lik	0.014	0.189	0.201	1000	0.033	0.138	0.141	1000
		pseudo-lik	0.010	0.190	0.202	1000	0.028	0.139	0.141	1000
	μ_ζ	lik	0.003	0.297	0.321	1000	-0.004	0.218	0.220	1000
		pseudo-lik	0.004	0.298	0.322	1000	-0.003	0.220	0.222	1000
	τ^2	lik	-0.434	0.346	0.401	999	-0.298	0.309	0.324	1000
		pseudo-lik	-0.432	0.348	0.406	998	-0.297	0.311	0.328	1000
	σ_ξ^2	lik	-0.061	0.179	0.191	1000	-0.072	0.130	0.137	999
		pseudo-lik	-0.059	0.179	0.191	999	-0.069	0.132	0.138	999
	σ_ζ^2	lik	-0.095	0.424	0.440	1000	-0.056	0.317	0.321	1000
		pseudo-lik	-0.085	0.428	0.448	1000	-0.042	0.321	0.324	1000

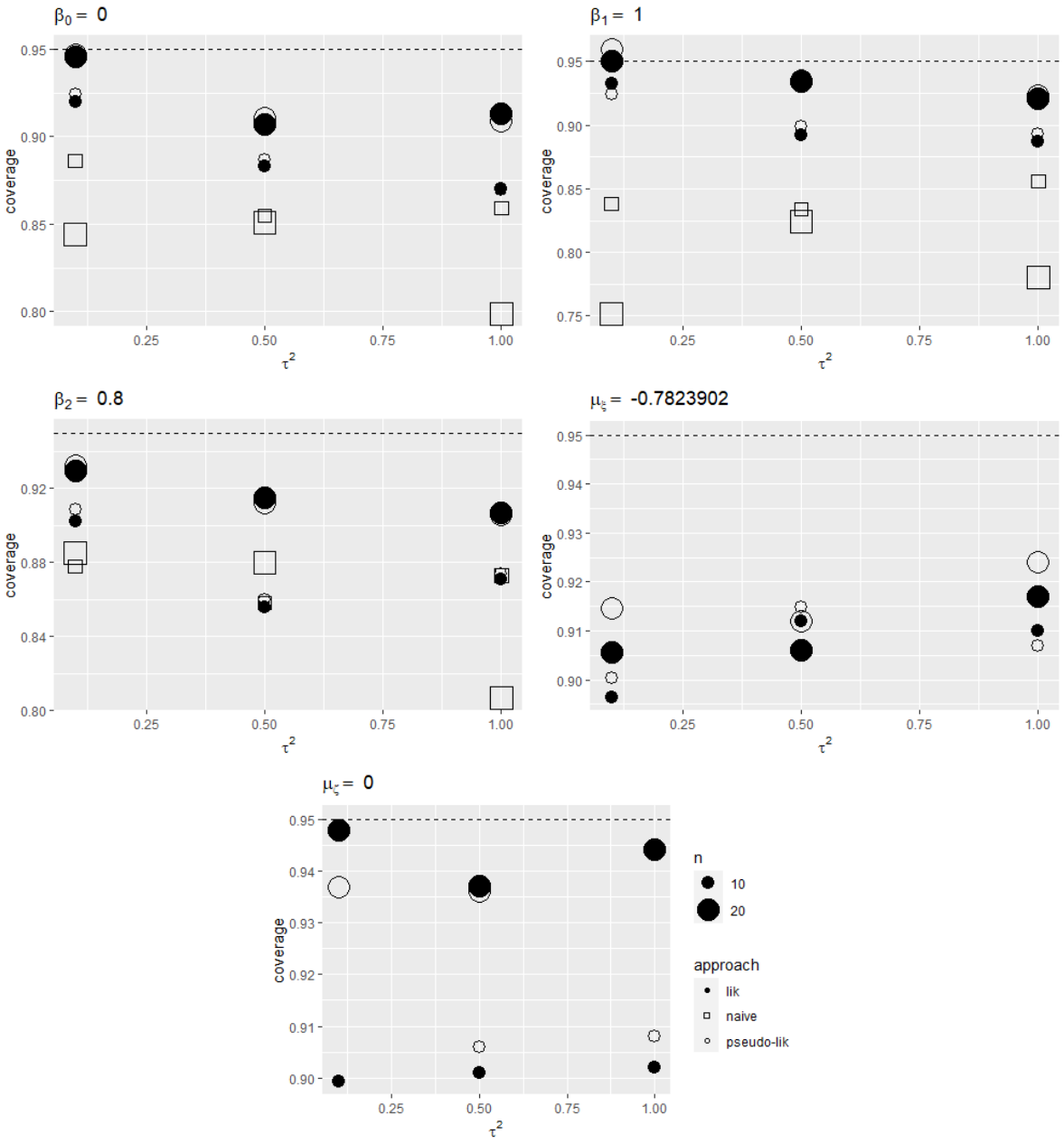

FIGURE A.4: Empirical coverage probabilities of 95% Wald-type confidence intervals for $(\beta_0, \beta_1, \beta_2, \mu_\xi, \mu_\varsigma)^\top$ for the uncorrected approach, the likelihood approach and the pseudo-likelihood approach when data follows scenario 3. Underlying risk distributed as a skew normal $SN(0, 1, -5)$.

Appendix

R code in myocardial injury example

```
rm(list=ls())
set.seed(1)
library(mvtnorm)
library(nlme)
library(metafor)
library(meta)
library(dplyr)
library(statmod)
setwd("C:/Users/trant/Dropbox/Thien_Phuc_Tran/software/simulation")

## functions
basic.lik.approx <- function(theta, dati, center=TRUE){
n <- nrow(dati)
beta0 <- theta[1]
beta1 <- theta[2]
mu.xi <- theta[3]
sigma2.eta <- theta[4]
sigma2.xi <-theta[5]
if(theta[4] < 0 | theta[5] < 0){ ## check variances
return(NA)
}
mean.vector <- c(beta0, mu.xi)
if(center==FALSE){mean.vector <- c(beta0+beta1*mu.xi, mu.xi)}
lik <- 0.0
for(i in 1:n){
```

```
S.matrix <- matrix(c(dati[i,3]+(beta1^2)*sigma2.xi+sigma2.eta, dati[i,4]+
beta1*sigma2.xi, dati[i,5]+beta1*sigma2.xi, dati[i,6]+sigma2.xi),
ncol=2, byrow=TRUE) ## S.matrix depends on i
lik <- lik + dmvnorm(dati[i,1:2], mean = mean.vector, sigma = S.matrix, log=TRUE)
}
return(lik)
}

basic.lik.GH3 <- function(theta, dati, n.node=10, model="approx", center=TRUE){
n <- nrow(dati)
beta0 <- theta[1]
beta1 <- theta[2]
mu.xi <- theta[3]
sigma2.eta <- theta[4]
sigma2.xi <-theta[5]
if(theta[4] < 0 | theta[5] < 0){ ## check variances
return(NA)
}
lik <- 0.0
## nodes and weights for Gauss-Hermite quadrature
objGH <- gauss.quad(n.node, "hermite")
w <- objGH$weights
w <- cbind(rep(w,rep(n.node,n.node)),rep(w,n.node))
w <- w[,1]*w[,2]
node <- objGH$nodes
node <- cbind(rep(node,rep(n.node,n.node)),rep(node,n.node))
if (model == "exact"){
for(i in 1:n){
g<-function(x){
xi <- sqrt(2*dati[i,6])*x[2]+dati[i,2]
eta <- sqrt(2*dati[i,3])*x[1]+dati[i,1]
p.eta <- exp(eta)/(1+exp(eta))
p.xi <- exp(xi)/(1+exp(xi))
return(dbinom(dati[i,7],dati[i,9],p.eta)*
dbinom(dati[i,8],dati[i,10],p.xi)*
dnorm(eta,beta0+beta1*(xi-mu.xi),sqrt(sigma2.eta))*
```

```
dnorm(xi,mu.xi,sqrt(sigma2.xi))*exp(x[1]^2+x[2]^2))
}
h <- apply(node, MARGIN=1, FUN=g)
lik <- lik + log(sum(w*h))
}
}
if(center==FALSE){
if (model == "exact"){
for(i in 1:n){
g<-function(x){
xi <- sqrt(2*dati[i,6])*x[2]+dati[i,2]
eta <- sqrt(2*dati[i,3])*x[1]+dati[i,1]
p.eta <- exp(eta)/(1+exp(eta))
p.xi <- exp(xi)/(1+exp(xi))
return(dbinom(dati[i,7],dati[i,9],p.eta)*
dbinom(dati[i,8],dati[i,10],p.xi)*
dnorm(eta,beta0+beta1*(xi),sqrt(sigma2.eta))*
dnorm(xi,mu.xi,sqrt(sigma2.xi))*exp(x[1]^2+x[2]^2))
}
h <- apply(node, MARGIN=1, FUN=g)
lik <- lik + log(sum(w*h))
}
}
}
return(lik)
}


quad.lik.GH3 <- function(theta, dati, n.node=10, model="approx"){
n <- nrow(dati)
beta0 <- theta[1]
beta1 <- theta[2]
beta2 <- theta[3]
mu.xi <- theta[4]
sigma2.eta <- theta[5]
sigma2.xi <-theta[6]
```

```
if(theta[5] < 0 | theta[6] < 0){ ## check variances
return(NA)
}
lik <- 0.0
## nodes and weights for Gauss-Hermite quadrature
objGH <- gauss.quad(n.node, "hermite")
w <- objGH$weights
w <- cbind(rep(w,rep(n.node,n.node)),rep(w,n.node))
w <- w[,1]*w[,2]
node <- objGH$nodes
node <- cbind(rep(node,rep(n.node,n.node)),rep(node,n.node))
if (model == "approx"){
for(i in 1:n){
g <- function(x){
xi <- sqrt(2*dati[i,7])*x[2]+dati[i,2]
eta <- sqrt(2*dati[i,4])*x[1]+dati[i,1]
return(dnorm(eta,beta0+beta1*(xi-mu.xi)+beta2*(xi-mu.xi)^2,sqrt(sigma2.eta))*
dnorm(xi,mu.xi,sqrt(sigma2.xi))/sqrt(pi*pi))
}
h <- apply(node, MARGIN=1, FUN=g)
lik <- lik + log(sum(w*h))
}
}
if (model == "exact"){
for(i in 1:n){
g<-function(x){
xi <- sqrt(2*dati[i,7])*x[2]+dati[i,2]
eta <- sqrt(2*dati[i,4])*x[1]+dati[i,1]
p.eta <- exp(eta)/(1+exp(eta))
p.xi <- exp(xi)/(1+exp(xi))
return(dbinom(dati[i,8],dati[i,10],p.eta)*dbinom(dati[i,9],dati[i,11],p.xi)*
dnorm(eta,beta0+beta1*(xi-mu.xi)+beta2*(xi-mu.xi)^2,sqrt(sigma2.eta))*
dnorm(xi,mu.xi,sqrt(sigma2.xi))*exp(x[1]^2+x[2]^2)*sqrt(2*dati[i,4]*2*dati[i,7]))
}
h <- apply(node, MARGIN=1, FUN=g)
lik <- lik + log(sum(w*h))
```

```
}
}
return(lik)
}

error.affect.lik.approx.repa <- function(theta, dati, pseudo=TRUE){
n <- nrow(dati)
beta0 <- theta[1]
beta1 <- theta[2]
beta2 <- theta[3]
mu.xi <- theta[4]
log.sigma2.eta <- theta[6]
log.sigma2.xi <-theta[7]

mu.z <- theta[5]
log.sigma2.z <- theta[8]

lik <- 0.0
mean.vector <- c(beta0+beta1*mu.xi+beta2*mu.z, mu.xi, mu.z)
#mean.vector <- c(beta0, mu.xi, mu.z)

for(i in 1:n){
if(pseudo==TRUE)
S.matrix <- matrix(c(dati[i,4]+(beta1^2)*exp(log.sigma2.xi)+(beta2^2)*
exp(log.sigma2.z)+exp(log.sigma2.eta),
beta1*exp(log.sigma2.xi), beta2*exp(log.sigma2.z),
beta1*exp(log.sigma2.xi), dati[i,8]+exp(log.sigma2.xi), 0,
beta2*exp(log.sigma2.z), 0, exp(log.sigma2.z)+dati[i,12]),
ncol=3, byrow=TRUE) ## S.matrix depends on i
else
S.matrix <- matrix(c(dati[i,4]+(beta1^2)*exp(log.sigma2.xi)+(beta2^2)
*exp(log.sigma2.z)+exp(log.sigma2.eta),
dati[i,5]+beta1*exp(log.sigma2.xi), dati[i,6]+beta2*exp(log.sigma2.z),
dati[i,7]+beta1*exp(log.sigma2.xi), dati[i,8]+exp(log.sigma2.xi), dati[i,9],
dati[i,10]+beta2*exp(log.sigma2.z), dati[i,11], exp(log.sigma2.z)+dati[i,12]),
ncol=3, byrow=TRUE) ## S.matrix depends on i
```

```
lik <- lik + dmvnorm(dati[i,1:3], mean = mean.vector, sigma = S.matrix, log=TRUE)
}
return(lik)
}

error.affect.lik.approx.repa.cen <- function(theta, dati, pseudo=TRUE){
n <- nrow(dati)
beta0 <- theta[1]
beta1 <- theta[2]
beta2 <- theta[3]
mu.xi <- theta[4]
log.sigma2.eta <- theta[6]
log.sigma2.xi <-theta[7]

mu.z <- theta[5]
log.sigma2.z <- theta[8]

lik <- 0.0
#mean.vector <- c(beta0+beta1*mu.xi+beta2*mu.z, mu.xi, mu.z)
mean.vector <- c(beta0, mu.xi, mu.z)

for(i in 1:n){
if(pseudo==TRUE)
S.matrix <- matrix(c(dati[i,4]+(beta1^2)*exp(log.sigma2.xi)+(beta2^2)*
exp(log.sigma2.z)+exp(log.sigma2.eta),
beta1*exp(log.sigma2.xi), beta2*exp(log.sigma2.z),
beta1*exp(log.sigma2.xi), dati[i,8]+exp(log.sigma2.xi), 0,
beta2*exp(log.sigma2.z), 0, exp(log.sigma2.z)+dati[i,12]), ncol=3, byrow=TRUE) ## S.matr
else
S.matrix <- matrix(c(dati[i,4]+(beta1^2)*exp(log.sigma2.xi)+(beta2^2)*
exp(log.sigma2.z)+exp(log.sigma2.eta),
dati[i,5]+beta1*exp(log.sigma2.xi), dati[i,6]+beta2*exp(log.sigma2.z),
dati[i,7]+beta1*exp(log.sigma2.xi), dati[i,8]+exp(log.sigma2.xi), dati[i,9],
dati[i,10]+beta2*exp(log.sigma2.z), dati[i,11], exp(log.sigma2.z)+dati[i,12]), ncol=3,
byrow=TRUE) ## S.matrix depends on i
```

```
lik <- lik + dmvnorm(dati[i,1:3], mean = mean.vector, sigma = S.matrix, log=TRUE)
}
return(lik)
}

## sanchez et al., 2021
y.i <- c(68,13,50,12,31,14,46,504,48,51,121,0,3,23)
x.i <- c(15,28,18,16,12,8,26,302,15,11,65,1,0,1)
ni.t <- c(94,21,123,23,52,24,112,914,89,133,170,10,16,50)
ni.c <- c(109,158,209,105,135,30,112,1906,35,538,989,125,85,95)
ni<-ni.t+ni.c

## create dataframe to generate forest plot
dat.fplot <- data.frame(tpos=y.i, tneg=ni.t-y.i, cpos=x.i, cneg=ni.c-x.i)

## calculate log risk ratios and corresponding sampling variances
dat.fplot <- escalc(measure="OR", ai=tpos, bi=tneg, ci=cpos, di=cneg,
data=dat.fplot, slab=c("Chen et al",
"Du et al", "Ferrante et al", "Franks et al",
"Guo et al", "Hingwei et al", "Lorente-Ros et al", "Mikami et al",
"Pan et al", "Shi et al", "Si et al", "Wan et al", "Wei et al", "Zhou et al"))

## fit random-effects model
res <- rma(yi, vi, data=dat.fplot)

## forest plot with extra annotations
forest(res)

#########################################################################
eta.hat <- log((y.i)/(ni.t-y.i))
w.new <- 1/( 1/(y.i)+1/(ni.t-y.i))
id.eta <- which(eta.hat==Inf | eta.hat==-Inf)
if(length(id.eta)>0){
eta.hat[id.eta] <- log((y.i[id.eta]+0.5)/(ni.t[id.eta]-y.i[id.eta]+0.5))
w.new[id.eta] <- 1/(1/(y.i[id.eta]+0.5)+1/(ni.t[id.eta]-y.i[id.eta]+0.5))
}
```

```
xi.hat <- log((x.i)/(ni.c-x.i))
id.xi <- which(xi.hat==Inf | xi.hat==-Inf)
if(length(id.xi)>0)
xi.hat[id.xi] <- log((x.i[id.xi]+0.5)/(ni.c[id.xi]-x.i[id.xi]+0.5))
var.eta <- 1/y.i+1/(ni.t-y.i)
var.xi <- 1/x.i+1/(ni.c-x.i)
id.eta <- which(var.eta==Inf)
if(length(id.eta)>0)
var.eta[id.eta] <- 1/(y.i[id.eta]+0.5)+1/(ni.t[id.eta]-y.i[id.eta]+0.5)
id.xi <- which(var.xi==Inf)
if(length(id.xi)>0)
var.xi[id.xi] <- 1/(x.i[id.xi]+0.5)+1/(ni.c[id.xi]-x.i[id.xi]+0.5)

######################################################################
## fit basic crr model to the dataset
dati <- data.frame(eta.hat=eta.hat, xi.hat=xi.hat, var.eta=var.eta,
cov.etaxi=0, cov.xieta=0, var.xi=var.xi, y.i=y.i, x.i=x.i, ni.t=ni.t,
ni.c=ni.c, ni=ni)
colnames(dati) <- c('y.obs', 'x.obs', 'var.y', 'cov.yx',
'cov.yx', 'var.x', 'y', 'x', 'ni.t', 'ni.c','ni')

## l'abbe plot
m <- metabin(y.i, ni.t, x.i, ni.c, data=dati, sm="OR", method="I",
backtransf=FALSE)
labbe(m, xlab="Control risk (log odds scale)", ylab="Treatment risk (log odds scale)")
abline(lm(y.obs ~ x.obs, data=dati, weights=1/dati$var.y)) ## WLS
abline(0, 1, lwd=2)

dati2 <- as_tibble(dati[,c(1,2,10)])
ggplot(dati2, aes(x=x.obs, y=y.obs, size=ni)) + geom_point() +
labs(y="treatment risk", x="control risk") + theme(legend.position="none")

n <- nrow(dati)
m.hat <- lm(y.obs ~ x.obs, data=dati) ## LS
est.naive <- c(coef(m.hat), mean(resid(m.hat)^2))
se.naive <- c(sqrt(diag(vcov(m.hat))), (sqrt( 2*(n-2)*est.naive[3]^2/n^2)))
```

```
theta.start <- c(coef(m.hat),                ## beta0, beta1
mean(dati[,2]),              ## mux
log(mean(resid(m.hat)^2)),   ##  (tau^2)
log(sd(dati[,2])^2))       ##  (sigmax^2)
model.approx <- optim(theta.start, basic.lik.approx.repa, dati=dati,
center=FALSE, control=list(fnscale=-1, maxit=5000), hessian=TRUE)
est <- model.approx$par
invH <- solve(model.approx$hessian)
se <- sqrt(diag(-diag(c(rep(1,3),exp(est[4:5])))%*%invH%*%diag(c(rep(1,3),
exp(est[4:5])))))
est[4:5]<-exp(est[4:5])

round(est, 3)
round(se, 3)
round(est.naive, 3)
round(se.naive,3)

##############################################################################
## fit crr with Z as age
zi.obs <- c(62,58,67,64,59,68,67,59,68,63,NA,47,49,56)

dati <- data.frame(eta.hat=eta.hat, xi.hat=xi.hat, zi.obs=zi.obs,
var.eta=var.eta, cov.etaxi=0, cov.etaz=0, cov.xieta=0, var.xi=var.xi, cov.xiz=0,
cov.zeta=0, cov.zxi=0, var.z=0, y.i=y.i, x.i=x.i, zi.obs=zi.obs, ni.t=ni.t,
ni.c=ni.c, ni=ni)
colnames(dati) <- c('y.obs', 'x.obs', 'zi.obs', 'var.y', 'cov.yx', 'cov.yz',
'cov.yx', 'var.x', 'cov.xz', 'cov.yz', 'cov.xz', 'var.z', 'y', 'x', 'zi', 'ni.t',
'ni.c', 'ni')
dati <- na.omit(dati) ## remove NAs

dati2 <- as_tibble(dati[,c(1,3,18)])
ggplot(dati2, aes(x=zi.obs, y=y.obs, size=ni)) + geom_point() +
labs(y="treatment risk", x="mean age") + theme(legend.position="none")

dati$zi.obs <- scale(dati$zi.obs, center=TRUE)
n <- nrow(dati)
```

```
m.hat <- lm(y.obs ~ x.obs + zi.obs, data=dati) ## LS
coef.naive <- c(coef(m.hat),                 ## beta0, beta1, beta2
mean(dati[,2]),               ## mu.xi
mean(dati[,3]),               ## mu.z
(mean(resid(m.hat)^2)))   ## (tau^2)
se.naive <- c(sqrt(diag(vcov(m.hat))), sd(dati[,2])^2/n, sd(dati[,3])^2/n,
(sqrt( 2*(n-3)*coef.naive[6]^2/n^2)))

theta.start <- c(coef(m.hat),                 ## beta0, beta1, beta2
mean(dati[,2]),               ## mux
mean(dati[,3]),               ## muz
log((mean(resid(m.hat)^2))),   ##  (tau^2)
log(sd(dati[,2])^2),        ##  (sigmax^2)
log(sd(dati[,3])^2))       ##  (sigmaz^2)

model.approx <- optim(theta.start, error.affect.lik.approx.repa, dati=dati,
pseudo=TRUE, control=list(fnscale=-1, maxit=5000),hessian=TRUE)
est<-model.approx$par
invH <- solve(model.approx$hessian)
se <- sqrt(diag(-diag(c(rep(1,5),exp(est[6:8])))%*%invH%*%diag(c(rep(1,5),
exp(est[6:8])))))
G <- matrix(0,length(est),length(est))
for(i in 1:nrow(dati)){
a <- fdHess(est, error.affect.lik.approx.repa, dati=dati[i,], pseudo=TRUE) ## compute He
values.gradient <- a$gradient
G <- G + values.gradient%*%t(values.gradient) ## compute J matrix
}
sand.se <- sqrt(diag(diag(c(rep(1,5),exp(est[6:8])))%*%invH%*%G%*%invH%*%
diag(c(rep(1,5),exp(est[6:8])))))
est[6:8] <- exp(est[6:8])

round(est, 3)
round(se, 3)
round(sand.se, 3)
round(coef.naive, 3)
round(se.naive, 3)
```

```
##############################################################################
## fit crr with Z as male
p.z <- c(62,54,71,57,49,NA,57,54,69,48,NA,53,54,62)/100 ##male
z.i <- p.z*ni
zi.obs <- log( (z.i)/(ni-z.i) )
id.zi <- which(zi.obs==Inf | zi.obs==-Inf)
if(length(id.zi)>0) ## check for infinite zi.obs and correct them
zi.obs[id.zi] <- log( (z.i[id.zi]+0.5)/(ni[id.zi]-z.i[id.zi]+0.5) )
var.z <- 1/z.i+1/(ni-z.i)
id.zi <- which(var.z==Inf)
if(length(id.zi)>0) ## check for infinite zi.obs variance and correct them
var.z[id.zi] <- 1/(z.i[id.zi]+0.5)+1/(ni[id.zi]-z.i[id.zi]+0.5)
dati <- data.frame(eta.hat=eta.hat, xi.hat=xi.hat, zi.obs=zi.obs, var.eta=var.eta,
cov.etaxi=0, cov.etaz=0, cov.xieta=0, var.xi=var.xi, cov.xiz=0,
cov.zeta=0, cov.zxi=0, var.z=var.z, y.i=y.i, x.i=x.i, ni=ni)
colnames(dati) <- c('y.obs', 'x.obs', 'zi.obs', 'var.y', 'cov.yx', 'cov.yz',
'cov.yx', 'var.x', 'cov.xz', 'cov.yz', 'cov.xz', 'var.z', 'y', 'x', 'ni')
dati <- na.omit(dati) ## remove NAs

dati2 <- as_tibble(dati[,c(1,3,15)])
ggplot(dati2, aes(x=zi.obs, y=y.obs, size=ni)) + geom_point() +
labs(y="treatment risk", x="log odds of male") + theme(legend.position="none")

n <- nrow(dati)
m.hat <- lm(y.obs ~ x.obs + zi.obs, data=dati) ## LS
coef.naive <- c(coef(m.hat),                  ## beta0, beta1, beta2
mean(dati[,2]),               ## mu.xi
mean(dati[,3]),               ## mu.z
(mean(resid(m.hat)^2)))   ## (tau^2)
se.naive <- c(sqrt(diag(vcov(m.hat))), sd(dati[,2])^2/n,
sd(dati[,3])^2/n, (sqrt(2*(n-3)*coef.naive[6]^2/n^2)))

theta.start <- c(coef(m.hat),                 ## beta0, beta1, beta2
mean(dati[,2]),               ## mux
mean(dati[,3]),               ## muz
log((mean(resid(m.hat)^2))),   ##  (tau^2)
```

```
log(sd(dati[,2])^2),        ##  (sigmax^2)
log(sd(dati[,3])^2))       ##  (sigmaz^2)

model.approx <- optim(theta.start, error.affect.lik.approx.repa, dati=dati,
pseudo=TRUE, control=list(fnscale=-1, maxit=5000),hessian=TRUE)
est <- model.approx$par
invH <- solve(model.approx$hessian)
se <- sqrt(diag(-diag(c(rep(1,5),exp(est[6:8])))%*%invH%*%diag(c(rep(1,5),
exp(est[6:8])))))
G <- matrix(0, length(est), length(est))
for(i in 1:nrow(dati)){
a <- fdHess(est, error.affect.lik.approx.repa, dati=dati[i,], pseudo=TRUE) ## compute He
values.gradient <- a$gradient
G <- G + values.gradient%*%t(values.gradient) ## compute J matrix
}
sand.se <- sqrt(diag(diag(c(rep(1,5),exp(est[6:8])))%*%invH%*%G%*%invH%*%
diag(c(rep(1,5),exp(est[6:8])))))
est[6:8] <- exp(est[6:8])

round(est, 3)
round(se, 3)
round(sand.se, 3)
round(coef.naive, 3)
round(se.naive, 3)

##############################################################################
## fit crr with Z as hypertension
p.z <- c(34,32,54,NA,33,24,75,25,50,30,NA,10,21,30)/100 ##hypertension

z.i <- p.z*ni
zi.obs <- log( (z.i)/(ni-z.i) )
id.zi <- which(zi.obs==Inf | zi.obs==-Inf)
if(length(id.zi)>0) ## check for infinite zi.obs and correct them
zi.obs[id.zi] <- log( (z.i[id.zi]+0.5)/(ni[id.zi]-z.i[id.zi]+0.5) )
var.z <- 1/z.i+1/(ni-z.i)
id.zi <- which(var.z==Inf)
```

```
if(length(id.zi)>0) ## check for infinite zi.obs variance and correct them
var.z[id.zi] <- 1/(z.i[id.zi]+0.5)+1/(ni[id.zi]-z.i[id.zi]+0.5)
dati <- data.frame(eta.hat=eta.hat, xi.hat=xi.hat, zi.obs=zi.obs,
var.eta=var.eta, cov.etaxi=0, cov.etaz=0, cov.xieta=0, var.xi=var.xi, cov.xiz=0,
cov.zeta=0, cov.zxi=0, var.z=var.z, y.i=y.i, x.i=x.i, ni=ni)
colnames(dati) <- c('y.obs', 'x.obs', 'zi.obs', 'var.y', 'cov.yx', 'cov.yz',
'cov.yx', 'var.x', 'cov.xz', 'cov.yz', 'cov.xz', 'var.z', 'y', 'x', 'ni')
dati <- na.omit(dati) ## remove NAs

dati2 <- as_tibble(dati[,c(1,3,15)])
ggplot(dati2, aes(x=zi.obs, y=y.obs, size=ni)) + geom_point() +
labs(y="treatment risk", x="log odds of hypertension")

n <- nrow(dati)
m.hat <- lm(y.obs ~ x.obs + zi.obs, data=dati) ## LS
coef.naive <- c(coef(m.hat),                  ## beta0, beta1, beta2
mean(dati[,2]),               ## mu.xi
mean(dati[,3]),               ## mu.z
(mean(resid(m.hat)^2)))   ## (tau^2)
se.naive <- c(sqrt(diag(vcov(m.hat))), sd(dati[,2])^2/n, sd(dati[,3])^2/n,
(sqrt(2*(n-3)*coef.naive[6]^2/n^2)))

theta.start <- c(coef(m.hat),                  ## beta0, beta1, beta2
mean(dati[,2]),               ## mux
mean(dati[,3]),               ## muz
log((mean(resid(m.hat)^2))),   ##  (tau^2)
log(sd(dati[,2])^2),        ##  (sigmax^2)
log(sd(dati[,3])^2))       ##  (sigmaz^2)

model.approx <- optim(theta.start, error.affect.lik.approx.repa, dati=dati,
pseudo=TRUE, control=list(fnscale=-1, maxit=5000),hessian=TRUE)
est<-model.approx$par
invH <- solve(model.approx$hessian)
se <- sqrt(diag(-diag(c(rep(1,5),exp(est[6:8])))%*%invH%*%
diag(c(rep(1,5),exp(est[6:8])))))
G <- matrix(0, length(est), length(est))
```

```
for(i in 1:nrow(dati)){
a <- fdHess(est, error.affect.lik.approx.repa, dati=dati[i,], pseudo=TRUE) ## compute He
values.gradient <- a$gradient
G <- G + values.gradient%*%t(values.gradient) ## compute J matrix
}
sand.se <- sqrt(diag(diag(c(rep(1,5),exp(est[6:8])))%*%invH%*%G%*%invH%*%
diag(c(rep(1,5),exp(est[6:8])))))
est[6:8] <- exp(est[6:8])

round(est, 3)
round(se, 3)
round(sand.se, 3)
round(coef.naive, 3)
round(se.naive, 3)

#########################################################################
## fit linear crr model
dati.basic <- data.frame(eta.hat=eta.hat, xi.hat=xi.hat, var.eta=var.eta,
cov.etaxi=0, cov.xieta=0, var.xi=var.xi, y.i=y.i, x.i=x.i, ni.t=ni.t, ni.c=ni.c)
colnames(dati.basic) <- c('y.obs', 'x.obs', 'var.y', 'cov.yx',
'cov.yx', 'var.x', 'y', 'x', 'ni.t', 'ni.c')
dati.basic$x.obs.cen <- scale(dati.basic$x.obs, scale=FALSE)
m <- nrow(dati.basic)

## naive estimation
m.hat.basic <- lm(y.obs ~ x.obs.cen, data=dati.basic)
naive.est.basic <- c(coef(m.hat.basic), mean(resid(m.hat.basic)^2))
naive.se.basic <- c(sqrt(diag(vcov(m.hat.basic))), sd(dati.basic$x.obs)/sqrt(m),
(sqrt(2*(m-2)*naive.est.basic[3]^2/m^2)), var(dati.basic$x.obs)*sqrt(2/(m-1)))

theta.start.basic <- c(coef(m.hat.basic),                ## beta0, beta1
mean(dati.basic$x.obs),               ## mu.xi
mean(resid(m.hat.basic)^2),   ##  tau2
sd(dati.basic$x.obs)^2)       ##  sigma2.xi
round(theta.start.basic, 3) ## estimate
round(naive.se.basic, 3)
```

```
round(theta.start.basic+1.96*naive.se.basic, 3) ## 95% wald-type confidence interv
round(theta.start.basic-1.96*naive.se.basic, 3) ## 95% wald-type confidence interv

## linear approximate model
model.approx.basic <- optim(theta.start.basic, basic.lik.approx, dati=dati.basic,
center=TRUE, control=list(fnscale=-1, maxit=5000), hessian=TRUE)
round(model.approx.basic$par,3) ## estimate

round(sqrt(diag(solve(-model.approx.basic$hessian))),3) ## hessian standard error
round(model.approx.basic$par+1.96*sqrt(diag(solve(-model.approx.basic$hessian))),
3) ## 95% wald-type confidence interval upper bound
round(model.approx.basic$par-1.96*sqrt(diag(solve(-model.approx.basic$hessian))),
3) ## 95% wald-type confidence interval lower bound
lal.max.lik<-model.approx.basic$value
round(2*5-lal.max.lik, 3)

## linear exact model
model.exactGH.basic <- optim(theta.start.basic, basic.lik.GH3, model="exact",
center=TRUE, dati=dati.basic, n.node=20, control=list(fnscale=-1, maxit=5000),
hessian=TRUE)
round(model.exactGH.basic$par, 3) ## estimate

round(sqrt(diag(solve(-model.exactGH.basic$hessian))), 3) ## hessian standard erro
round(model.exactGH.basic$par+1.96*sqrt(diag(solve(-model.exactGH.basic$hessian)))
round(model.exactGH.basic$par-1.96*sqrt(diag(solve(-model.exactGH.basic$hessian)))
lel.max.lik<-model.exactGH.basic$value
lel.max.lik<-lel.max.lik+sum(log(sqrt(2*dati.basic[,3]*2*dati.basic[,6])))
round(2*5-lel.max.lik, 3)

## fit quadratic crr model
dati <- data.frame(eta.hat=eta.hat, xi.hat=xi.hat, xi.hat2=xi.hat^2,
var.eta=var.eta, cov.etaxi=0, cov.xieta=0, var.xi=var.xi, y.i=y.i, x.i=x.i,
ni.t=ni.t, ni.c=ni.c)
colnames(dati) <- c('y.obs', 'x.obs', 'x.obs2', 'var.y', 'cov.yx',
'cov.yx', 'var.x', 'y', 'x', 'ni.t', 'ni.c')
dati$x.obs.cen <- scale(dati$x.obs, scale=FALSE)
```

```
dati$x.obs.cen2 <- dati$x.obs.cen^2

## naive estimation
m.hat <- lm(y.obs ~ x.obs.cen+x.obs.cen2, data=dati)
naive.est <- c(coef(m.hat), mean(resid(m.hat)^2))

theta.start <- c(coef(m.hat),                 ## beta0, beta1, beta2
mean(dati$x.obs),               ## mu.xi
mean(resid(m.hat)^2),   ## tau2
sd(dati$x.obs)^2)       ## sigma2.xi

round(theta.start, 3) ## estimate
naive.se <- c(sqrt(diag(vcov(m.hat))), sd(dati$x.obs)/sqrt(m),
(sqrt(2*(m-3)*naive.est[4]^2/m^2)), var(dati$x.obs)*sqrt(2/(m-1))) ## standard error
round(naive.se, 3)
round(theta.start+1.96*naive.se, 3) ## 95% wald-type confidence interval upper bound
round(theta.start-1.96*naive.se, 3) ## 95% wald-type confidence interval lower bound

## quadratic approximate model
model.approxGH <- optim(theta.start, quad.lik.GH3, dati=dati, n.node=20,
model='approx', control=list(fnscale=-1, maxit=5000), hessian=TRUE)
round(model.approxGH$par,3)

round(sqrt(diag(solve(-model.approxGH$hessian))),3) ## hessian standard error
round(model.approxGH$par+1.96*sqrt(diag(solve(-model.approxGH$hessian))),3) ## 95% wald-
round(model.approxGH$par-1.96*sqrt(diag(solve(-model.approxGH$hessian))),3) ## 95% wald-
qal.max.lik<-model.approxGH$value
qal.max.lik<-qal.max.lik-log(sqrt(pi*pi))
round(2*6-qal.max.lik, 3)

## quadratic exact model
model.exactGH <- optim(theta.start, quad.lik.GH3, dati=dati, n.node=20, model='exact',
control=list(fnscale=-1, maxit=5000), hessian=TRUE)
round(model.exactGH$par,3) ## estimate

round(sqrt(diag(solve(-model.exactGH$hessian))), 3) ## hessian standard error
```

```
round(model.exactGH$par+1.96*sqrt(diag(solve(-model.exactGH$hessian))), 3) ## 95%
round(model.exactGH$par-1.96*sqrt(diag(solve(-model.exactGH$hessian))), 3) ## 95%
qel.max.lik<-model.exactGH$value
qel.max.lik<-qel.max.lik+sum(log(sqrt(2*dati[,4]*2*dati[,7])))
round(2*6-qel.max.lik, 3)

## model visualization
plot(xi.hat, eta.hat, xlab="control risk", ylab="treatment risk")

## quadratic approximate model
lines(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05), model.approxGH$par[1]
+model.approxGH$par[2]*(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05)-
model.approxGH$par[4])
+model.approxGH$par[3]*(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05)-
model.approxGH$par[4])^2,
lwd = 2, lty = 1)

## quadratic exact model
lines(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05), model.exactGH$par[1]
+model.exactGH$par[2]*(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05)-
model.approxGH$par[4])
+model.exactGH$par[3]*(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05)-
model.approxGH$par[4])^2,
lwd = 2, lty = 2)

## quadratic naive model
lines(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05), theta.start[1]+
theta.start[2]*(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05)-theta.start[4])+
theta.start[3]*(seq(min(xi.hat)-1, max(xi.hat)+1, 0.05)-theta.start[4])^2,
lwd = 2, lty = 3)

## linear approximate model
abline(a=model.approx.basic$par[1]-model.approx.basic$par[2]*
model.approx.basic$par[3], b=model.approx.basic$par[2], lwd = 1, lty=1)

## linear exact model
```

```
abline(a=model.exactGH.basic$par[1]-model.approx.basic$par[2]*
model.approx.basic$par[3], b=model.exactGH.basic$par[2], lwd = 1, lty = 2)

## linear regression model
abline(a=theta.start.basic[1]-theta.start.basic[2]*theta.start.basic[3],
b=theta.start.basic[2], lwd = 1, lty = 3)

## legend
legend("topleft", legend = c("QAL", "QEL", "QNA", "LAL", "LEL", "LNA"),
lwd = rep(c(2,1), c(3,3)), lty = c(1, 2, 3, 1, 2, 3), title='Method',
cex=0.5, ncol=2)

#########################################################################
## refit model with hypertension
p.z <- c(34,32,54,NA,33,24,75,25,50,30,NA,10,21,30)/100 ##hypertension

z.i <- p.z*ni

zi.obs <- log( (z.i)/(ni-z.i) )

id.zi <- which(zi.obs==Inf | zi.obs==-Inf)
if(length(id.zi)>0) ## check for infinite zi.obs and correct them
zi.obs[id.zi] <- log( (z.i[id.zi]+0.5)/(ni[id.zi]-z.i[id.zi]+0.5) )

var.z <- 1/z.i+1/(ni-z.i)
id.zi <- which(var.z==Inf)
if(length(id.zi)>0) ## check for infinite zi.obs variance and correct them
var.z[id.zi] <- 1/(z.i[id.zi]+0.5)+1/(ni[id.zi]-z.i[id.zi]+0.5)

dati <- data.frame(eta.hat=eta.hat, xi.hat=xi.hat, zi.obs=zi.obs, var.eta=var.eta,
cov.etaxi=0, cov.etaz=0, cov.xieta=0, var.xi=var.xi, cov.xiz=0,
cov.zeta=0, cov.zxi=0, var.z=var.z, y.i=y.i, x.i=x.i, ni=ni)

colnames(dati) <- c('y.obs', 'x.obs', 'zi.obs', 'var.y', 'cov.yx', 'cov.yz',
'cov.yx', 'var.x', 'cov.xz', 'cov.yz', 'cov.xz', 'var.z', 'y', 'x', 'ni')
```

```
dati <- na.omit(dati) ## remove NAs
dati$x.obs.cen <- scale(dati$x.obs, scale=FALSE)
dati$zi.obs.cen <- scale(dati$zi.obs, scale=FALSE)

n <- nrow(dati)
m.hat <- lm(y.obs ~ x.obs.cen + zi.obs.cen, data=dati) ## LS
theta.start <- c( coef(m.hat),                 ## beta0, beta1, beta2
mean(dati[,2]),               ## mux
mean(dati[,3]),               ## muz
log((mean(resid(m.hat)^2))),   ##  (tau^2)
log(sd(dati[,2])^2),        ##  (sigmax^2)
log(sd(dati[,3])^2))       ##  (sigmaz^2)

model.approx <- optim(theta.start, error.affect.lik.approx.repa.cen, dati=dati,
pseudo=TRUE, control=list(fnscale=-1, maxit=5000),hessian=TRUE)
est<-model.approx$par
invH <- solve(model.approx$hessian)
se <- sqrt(diag(-diag(c(rep(1,5),exp(est[6:8])))%*%invH%*%diag(c(rep(1,5),
exp(est[6:8])))))
G <- matrix(0,length(est),length(est))
for(i in 1:nrow(dati)){
a <- fdHess(est, error.affect.lik.approx.repa, dati=dati[i,], pseudo=TRUE) ## comp
values.gradient <- a$gradient
G <- G + values.gradient%*%t(values.gradient) ## compute J matrix
}
sand.se <- sqrt(diag(diag(c(rep(1,5),exp(est[6:8])))%*%invH%*%G%*%invH%*%
diag(c(rep(1,5),exp(est[6:8])))))
est[6:8]<-exp(est[6:8])

round(est,3)
round(se,3)
round(sand.se,3)
round(2*8-2*model.approx$value,3)
```

Bibliography

Arends, L. R., Hoes, A. W., Lubsen, J., Grobbee, D. E. and Stijnen, T. (2000) Baseline risk as predictor of treatment benefit: Three clinical meta-re-analyses. *Statistics in Medicine* **19**(24), 3497–3518.

Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**(2), 171–178.

Bagnardi, V., Zambon, A., Quatto, P. and Corrao, G. (2004) Flexible meta-regression functions for modeling aggregate dose-response data, with an application to alcohol and mortality. *American Journal of Epidemiology* **159**(11), 1077–1086.

Bagos, P. G. (2012) On the covariance of two correlated log-odds ratios. *Statistics in Medicine* **31**(14), 1418–1431.

Baker, W. L., Michael White, C., Cappelleri, J. C., Kluger, J., Coleman, C. I., Health Outcomes, P. and Group, E. H. C. (2009) Understanding heterogeneity in meta-analysis: The role of meta-regression. *International Journal of Clinical Practice* **63**(10), 1426–1434.

Bellio, R. and Guolo, A. (2016) Integrated likelihood inference in small sample meta-analysis for continuous outcomes. *Scandinavian Journal of Statistics* **43**(1), 191–201.

Berkey, C. S., Hoaglin, D. C., Mosteller, F. and Colditz, G. A. (1995) A random-effects regression model for meta-analysis. *Statistics in Medicine* **14**(4), 395–411.

Berlin, J. A., Santanna, J., Schmid, C. H., Szczech, L. A. and Feldman, H. I. (2002) Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: Ecological bias rears its ugly head. *Statistics in Medicine* **21**(3), 371–387.

Bernsen, R. M. D., Tasche, M. J. A. and Nagelkerke, N. J. D. (1999) Variation in baseline risk as an explanation of heterogeneity in meta-analysis by s. d. walter, statistics in medicine, 16, 2883-2900 (1997). *Statistics in Medicine* **18**(2), 233–237.

Besag, J. (1975) Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society. Series D (The Statistician)* **24**(3), 179–195.

Böhning, D. (2000) *Computer-Assisted Analysis of Mixtures and Applications: Meta-Analysis, Disease Mapping and Others*. Third edition. Boca Raton: Chapman and Hall/CRC,.

Böhning, D. (2005) Meta-analysis: A unifying meta-likelihood approach framing unobserved heterogeneity, study covariates, publication bias, and study quality. *Methods of Information in Medicine* **44**(1), 127–135.

Bland, M. J. and Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* **327**(8476), 307–310.

Boissel, J., Cucherat, M., Nony, P., Chabaud, S., Gueyffier, F., Wright, J. M., Lièvre, M. and Leizorovicz, A. (2008) New insights on the relation between untreated and treated outcomes for a given therapy effect model is not necessarily linear. *Journal of Clinical Epidemiology* **61**(3), 301–307.

Borenstein, M., Hedges, L. V., Higgins, J. P. and Rothstein, H. R. (2010) A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods* **1**(2), 97–111.

Brand, R. and Kragt, H. (1992) Importance of trends in the interpretation of an overall odds ratio in the meta-analysis of clinical trials. *Statistics in Medicine* **11**(16), 2077–2082.

Brockwell, S. E. and Gordon, I. R. (2001) A comparison of statistical methods for meta-analysis. *Statistics in Medicine* **20**(6), 825–840.

Burr, D. and Doss, H. (2005) A Bayesian semiparametric model for random-effects meta-analysis. *Journal of the American Statistical Association* **100**(469), 242–251.

Burr, D., Doss, H., Cooke, G. E. and Goldschmidt-Clermont, P. J. (2003) A meta-analysis of studies on the association of the platelet pla polymorphism of glycoprotein iiia and risk of coronary heart disease. *Statistics in Medicine* **22**(10), 1741–1760.

Carroll, R., Ruppert, D., Stefanski, L. and Crainiceanu, C. (2006) *Measurement Error in Nonlinear Models: A Modern Perspective*. Second edition. Boca Raton: Chapman and Hall/CRC.

Cereda, E., Barichella, M., Pedrolli, C., Klersy, C., Cassani, E., Caccialanza, R. and Pezzoli, G. (2011) Diabetes and risk of Parkinson's disease: A systematic review and meta-analysis. *Diabetes Care* **34**(12), 2614–2623.

Cereda, E., Barichella, M., Pedrolli, C., Klersy, C., Cassani, E., Caccialanza, R. and Pezzoli, G. (2013) Diabetes and risk of parkinson's disease. *Movement Disorders* **28**(2), 257–261.

Chaimani, A. (2015) Accounting for baseline differences in meta-analysis. *BMJ Mental Health* **18**(1), 23–26.

Cochran, W. G. (1937) Problems arising in the analysis of a series of similar experiments. *Supplement to the Journal of the Royal Statistical Society* **4**(1), 102–118.

Cochran, W. G. (1954) The combination of estimates from different experiments. *Biometrics* **10**(1), 101–129.

Cook, J. R. and Stefanski, L. A. (1994) Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **89**(428), 1314–1328.

Cook, R. J. and Walter, S. D. (1997) A logistic model for trend in 2 x 2 x k tables with applications to meta-analyses. *Biometrics* **53**(1), 352–357.

Craft, S. and Watson, G. S. (2004) Insulin and neurodegenerative disease: Shared and specific mechanisms. *The Lancet Neurology* **3**(3), 169–178.

DerSimonian, R. and Kacker, R. (2007) Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials* **28**(2), 105–114.

DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**(3), 177–188.

Fernandez, C. and Steel, M. F. J. (1998) On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **93**(441), 359–371.

Follmann, D. A. and Proschan, M. A. (1999) Valid inference in random-effects meta-analysis. *Biometrics* **55**(3), 732–737.

Fond, G., Pauly, V., Leone, M., Llorca, P., Orleans, V., Loundou, A., Lancon, C., Auquier, P., Baumstarck, K. and Boyer, L. (2020) Disparities in intensive care unit admission and mortality among patients with schizophrenia and covid-19: A national cohort study. *Schizophrenia Bulletin* **47**(3), 624–634.

Gelman, A. B., Carlin, J. B., Stern, H. S. and Dunson, D. B. (2013) *Bayesian Data Analysis*. Third edition. Boca Raton: Chapman andHall/CRC.

Ghidey, W., Lesaffre, E. and Stijnen, T. (2007) Semi-parametric modelling of the distribution of the baseline risk in meta-analysis. *Statistics in Medicine* **26**(30), 5434–5444.

Ghidey, W., Stijnen, T. and van Houwelingen, H. C. (2013) Modelling the effect of baseline risk in meta-analysis: A review from the perspective of errors-in-variables regression. *Statistical Methods in Medical Research* **22**(3), 307–323.

Glass, G. V. and Smith, M. L. (1979) Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis* **1**(1), 2–16.

Glasziou, P. P. and Sanders, S. L. (2002) Investigating causes of heterogeneity in systematic reviews. *Statistics in Medicine* **21**(11), 1503–1511.

Gold, M. S., Sehayek, D., Gabrielli, S., Zhang, X., McCusker, C. and Ben-Shoshan, M. (2020) Covid-19 and comorbidities: A systematic review and meta-analysis. *Postgraduate Medicine* **132**(8), 749–755.

Guo, T., Fan, Y., Chen, M., Wu, X., Zhang, L., He, T., Wang, H., Wan, J., Wang, X. and Lu, Z. (2020) Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (covid-19). *JAMA Cardiology* **5**(7), 811–818.

Guolo, A. (2008) A flexible approach to measurement error correction in case-control studies. *Biometrics* **64**(4), 1207–1214.

Guolo, A. (2012) Higher-order likelihood inference in meta-analysis and meta-regression. *Statistics in Medicine* **31**(4), 313–327.

Guolo, A. (2013) Flexibly modeling the baseline risk in meta-analysis. *Statistics in Medicine* **32**(1), 40–50.

Guolo, A. (2014) The simex approach to measurement error correction in meta-analysis with baseline risk as covariate. *Statistics in Medicine* **33**(12), 2062–2076.

Guolo, A. (2021) Measurement error and misclassification in meta-analysis. In *Handbook of Mesurement Error Models*, eds G. Y. Yi, A. Delaigle and P. Gustafson. Boca Raton: Chapman and Hall/CRC, first edition.

Guolo, A. (2022) Measurement errors in control risk regression: A comparison of correction techniques. *Statistics in Medicine* **41**(1), 163–179.

Guolo, A., Schmid, C. H. and Stijnen, T. (2021) Control risk regression. In *Handbook of Meta-Analysis*, eds C. H. Schmid, T. Stijnen and I. R. White. Boca Raton: Chapman and Hall/CRC, first edition.

Guolo, A. and Varin, C. (2017) Random-effects meta-analysis: The number of studies matters. *Statistical Methods in Medical Research* **26**(3), 1500–1518.

Gurevitch, J., Koricheva, J., Nakagawa, S. and Stewart, G. (2018) Meta-analysis and the science of research synthesis. *Nature* **555**, 1476–4687.

Hamza, T. H., van Houwelingen, H. C. and Stijnen, T. (2008) The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of Clinical Epidemiology* **61**(1), 41–51.

Hardy, R. J. and Thompson, S. G. (1996) A likelihood approach to meta-analysis with random effects. *Statistics in Medicine* **15**(6), 619–629.

Hardy, R. J. and Thompson, S. G. (1998) Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine* **17**(8), 841–856.

Hariyanto, T. I. and Kurniawan, A. (2021) Obstructive sleep apnea (OSA) and outcomes from coronavirus disease 2019 (COVID-19) pneumonia: A systematic review and meta-analysis. *Sleep Medicine* **82**, 47–53.

Hartung, J. and Knapp, G. (2001a) A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Statistics in Medicine* **20**(24), 3875–3889.

Hartung, J. and Knapp, G. (2001b) On tests of the overall treatment effect in meta-analysis with normally distributed responses. *Statistics in Medicine* **20**(12), 1771–1782.

Hedges, L. V. and Olkin, I. (1985) *Statistical Methods for Meta-Analysis*. San Diego: Academic Press.

Hedges, L. V. and Pigott, T. D. (2004) The power of statistical tests for moderators in meta-analysis. *Psychological Methods* **9**(4), 426–445.

Higgins, J. P. and Thompson, S. G. (2004) Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine* **23**(11), 1663–1682.

Higgins, J. P. T. and Thompson, S. G. (2002) Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**(11), 1539–1558.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J. and Altman, D. G. (2003) Measuring inconsistency in meta-analyses. *British Medical Journal* **327**(7414), 557–560.

Higgins, J. P. T., Thompson, S. G. and Spiegelhalter, D. J. (2009) A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society* **172**(1), 137–159.

van Houwelingen, H. and Senn, S. (1999) Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* **18**(1), 110–115.

van Houwelingen, H. C., Arends, L. R. and Stijnen, T. (2002) Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine* **21**(4), 589–624.

Huizenga, H. M., Visser, I. and Dolan, C. V. (2011) Testing overall and moderator effects in random effects meta-regression. *British Journal of Mathematical and Statistical Psychology* **64**(1), 1–19.

Hunter, J. and Schmidt, F. (2015) *Methods of Meta-Analysis Corrected Error and Bias in Research Findings*. Third edition. Thousand Oaks: SAGE Publications, Ltd.

Hunter, J. E. and Schmidt, F. L. (2000) Fixed-effects vs. random-effects meta-analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment* **8**(4), 275–292.

Jackson, D. (2013) Confidence intervals for the between-study variance in random effects meta-analysis using generalised cochran heterogeneity statistics. *Research Synthesis Methods* **4**(3), 220–229.

Jackson, D., Bowden, J. and Baker, R. (2010) How does the dersimonian and laird procedure for random effects meta-analysis compare with its more efficient but harder to compute counterparts? *Journal of Statistical Planning and Inference* **140**(4), 961–970.

Kauermann, G. and Carroll, R. J. (2001) A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* **96**(456), 1387–1396.

Knapp, G., Biggerstaff, B. and Hartung, J. (2006) Assessing the amount of heterogeneity in random-effects meta-analysis. *Biometrical Journal* **48**(2), 271–285.

Knapp, G. and Hartung, J. (2003) Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* **22**(17), 2693–2710.

Koricheva, J., Gurevitch, J. and Mengersen, K. (2013) *The Handbook of Meta-Analysis in Ecology and Evolution*. New Jersey: Princeton University Press.

Kroken, R. A., Sommer, I. E., Steen, V. M., Dieset, I. and Johnsen, E. (2019) Constructing the immune signature of schizophrenia for clinical use and research; an integrative review translating descriptives into diagnostics. *Frontiers in Psychiatry* **9**, 753.

L'Abbé, K. A., Detsky, A. S. and O'Rourke, K. (1987) Meta-analysis in clinical research. *Annals of Internal Medicine* **107**(2), 224–233.

Laird, N. (1978) Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**(364), 805–811.

Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., Viechtbauer, W. and Simmonds, M. (2019) A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods* **10**(1), 83–98.

Lee, K. J. and Thompson, S. G. (2008) Flexible parametric models for random-effects distributions. *Statistics in Medicine* **27**(3), 418–434.

López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., Van den Noortgate, W. and Viechtbauer, W. (2014) Estimation of the predictive power of the model in mixed-effects meta-regression: A simulation study. *British Journal of Mathematical and Statistical Psychology* **67**(1), 30–48.

Lu, L., Fu, D. L., Li, H. Q., Liu, A. J., Li, J. H. and Zheng, G. Q. (2014) Diabetes and risk of parkinson's disease: An updated meta-analysis of case-control studies. *PloS One* **9**(1), e85781.

McIntosh, M. W. (1996) The population risk as an explanatory variable in research synthesis of clinical trials. *Statistics in Medicine* **15**(16), 1713–1728.

Morris, T. P., White, I. R. and Crowther, M. J. (2019) Using simulation studies to evaluate statistical methods. *Statistics in Medicine* **38**(11), 2074–2102.

Nelder, J. A. and Mead, R. (1965) A simplex method for function minimization. *The Computer Journal* **7**(4), 308–313.

Ohlssen, D. I., Sharples, L. D. and Spiegelhalter, D. J. (2007) Flexible random-effects models using Bayesian semi-parametric models: Applications to institutional comparisons. *Statistics in Medicine* **26**(9), 2088–2112.

Papadimitropoulou, K., Stijnen, T., Dekkers, O. and Le Cessie, S. (2019) One-stage random effects meta-analysis using linear mixed models for aggregate continuous outcome data. *Research Synthesis Methods* **10**(3), 360–375.

Pardamean, E., Roan, W., Iskandar, K. T. A., Prayangga, R. and Hariyanto, T. I. (2022) Mortality from coronavirus disease 2019 (Covid-19) in patients with schizophrenia: A systematic review, meta-analysis and meta-regression. *General Hospital Psychiatry* **75**, 61–67.

Paule, R. C. and Mandel, J. (1982) Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards* **87**(5), 377–385.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raudenbush, S. W. (1994) Random effects models. In *Handbook of Research Synthesis, The*, eds H. Cooper and L. V. Hedges. New York: Russell Sage Foundation.

Raudenbush, S. W. (2009) Analyzing effect sizes: Random-effects models. In *The Handbook of Research Synthesis and Meta-Analysis*, eds H. Cooper, L. V. Hedges and J. C. Valentine. New York: Russell Sage Foundation, second edition.

Raudenbush, S. W. and Bryk, A. S. (1985) Empirical Bayes meta-analysis. *Journal of Educational Statistics* **10**(2), 75–98.

Rhodes, K. M., Turner, R. M. and Higgins, J. P. (2015) Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *Journal of Clinical Epidemiology* **68**(1), 52–60.

Rice, K., Higgins, J. P. T. and Lumley, T. (2018) A re-evaluation of fixed effect(s) meta-analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society* **181**(1), 205–227.

Rivas-Ramírez, A. R., Tendilla-Beltrán, H., Gómez-Mendoza, L. E., Loaiza, G. and Flores, G. (2021) Patients with schizophrenia have decreased COVID-19 prevalence among hospitalised patients with psychiatric and neurological diseases: A retrospective analysis in Mexican population. *International Journal of Clinical Practice* **75**(10), e14528.

Rodrigues-Amorim, D., Rivera-Baltanás, T., Spuch, C., Caruncho, H. J., González-Fernandez, A., Olivares, J. M. and Agís-Balboa, R. C. (2018) Cytokines dysregulation in schizophrenia: A systematic review of psychoneuroimmune relationship. *Schizophrenia Research* **197**, 19–33.

Sanchez-Meca, J. and Marín-Martínez, F. (1997) Homogeneity tests in meta-analysis: A Monte Carlo comparison of statistical power and Type I error. *Quality and Quantity* **31**, 385–399.

Santiago, J. A. and Potashkin, J. A. (2013) Shared dysregulated pathways lead to parkinson's disease and diabetes. *Trends in Molecular Medicine* **19**(3), 176–186.

Sanz-Sánchez, J., Vrachatis, D. A., Reimers, B., Deftereos, S. G., Kallikourdis, M., Vicenzi, M., Giannopoulos, G., Giotaki, S. G., Tousoulis, D., Ferrante, G., Condorelli, G. and Stefanini, G. G. (2021) Impact of myocardial injury on mortality in patients with COVID-19: A meta-analysis. *Hellenic Journal of Cardiology* **62**(3), 253–255.

Schmid, C. H., Carlin, B. P. and Welton, N. J. (2021) Bayesian methods for meta-analysis. In *Handbook of Meta-Analysis*, eds C. H. Schmid, T. Stijnen and I. R. White. Boca Raton: Chapman andHall/CRC, first edition.

Schmid, C. H., Lau, J., McIntosh, M. W. and Cappelleri, J. C. (1998) An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Statistics in Medicine* **17**(17), 1923–1942.

Schmid, C. H., Stark, P. C., Berlin, J. A., Landais, P. and Lau, J. (2004) Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *Journal of Clinical Epidemiology* **57**(7).

Schmidt, F. and Hunter, J. (2015) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Third edition. SAGE Publication.

Schmidt, F. L., Oh, I. and Hayes, T. L. (2009) Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology* **62**(1), 97–128.

Schulze, R. (2004) *Meta-Analysis - A Comparison of Approaches*. First edition. Boston: Hogrefe & Huber Publishers.

Severini, T. A. (2000) *Likelihood Methods in Statistics*. Oxford: Oxford University Press.

Shadish, W. R. and Lecy, J. D. (2015) The meta-analytic big bang. *Research Synthesis Methods* **6**(3), 246–264.

Sharp, S. J. and Thompson, S. G. (2000) Analysing the relationship between treatment effect and underlying risk in meta-analysis: Comparison and development of approaches. *Statistics in Medicine* **19**(23), 3251–3274.

Sharp, S. J., Thompson, S. G. and Altman, D. G. (1996) The relation between treatment benefit and underlying risk in meta-analysis. *BMJ: British Medical Journal* **313**(7059), 735–738.

Shi, S., Qin, M., Shen, B., Cai, Y., Liu, T., Yang, F., Gong, W., Liu, X., Liang, J., Zhao, Q., Huang, H., Yang, B. and Huang, C. (2020) Association of cardiac injury with mortality in hospitalized patients with covid-19 in Wuhan, China. *JAMA Cardiology* **5**(7), 802–810.

Sidik, K. and Jonkman, J. N. (2002) A simple confidence interval for meta-analysis. *Statistics in Medicine* **21**(21), 3153–3159.

Sidik, K. and Jonkman, J. N. (2005a) A note on variance estimation in random effects meta-regression. *Journal of Biopharmaceutical Statistics* **15**(5), 823–838.

Sidik, K. and Jonkman, J. N. (2005b) Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **54**(2), 367–384.

Simmonds, M. C. and Higgins, J. P. T. (2007) Covariate heterogeneity in meta-analysis: Criteria for deciding between meta-regression and individual patient data. *Statistics in Medicine* **26**(15), 2982–2999.

Skovgaard, I. M. (1996) An explicit large-deviation approximation to one-parameter tests. *Bernoulli* **2**(2), 145–165.

Smith, T. C., Spiegelhalter, D. J. and Thomas, A. (1995) Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine* **14**(24), 2685–2699.

Spiegelhalter, D. J., Abrams, K. R. and Myles, J. P. (2003) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. First edition. New Jersey: John Wiley & Sons, Ltd.

Stefanski, L. A. and Carroll, R. J. (1987) Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74**(4), 703–716.

Stefanski, L. A. and Cook, J. R. (1995) Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association* **90**(432), 1247–1256.

Sterne, J. A. C. and Smith, G. D. (2001) Sifting the evidence-what's wrong with significance tests? *British Medical Journal* **322**(7280), 226–231.

Stijnen, T., Hamza, T. H. and Özdemir, P. (2010) Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine* **29**(29), 3046–3067.

Stijnen, T., White, I. R. and Schmid, C. H. (2021) Analysis of univariate study-level summary data using normal models. In *Handbook of Meta-Analysis*, eds C. H. Schmid, T. Stijnen and I. R. White. Boca Raton: Chapman and Hall/CRC, first edition.

Sutton, A. J. and Abrams, K. R. (2001) Bayesian methods in meta-analysis and evidence synthesis. *Statistical Methods in Medical Research* **10**(4), 277–303.

Sutton, A. J. and Higgins, J. P. T. (2008) Recent developments in meta-analysis. *Statistics in Medicine* **27**(5), 625–650.

Thompson, S. G. and Higgins, J. P. T. (2002) How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* **21**(11), 1559–1573.

Thompson, S. G. and Sharp, S. J. (1999) Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine* **18**(20), 2693–2708.

Thompson, S. G., Smith, T. C. and Sharp, S. J. (1997) Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statistics in Medicine* **16**(23), 2741–2758.

Tipton, E., Pustejovsky, J. E. and Ahmadi, H. (2019a) A history of meta-regression: Technical, conceptual, and practical developments between 1974 and 2018. *Research Synthesis Methods* **10**(2), 161–179.

Tipton, E., Pustejovsky, J. E. and Ahmadi, H. (2019b) Current practices in meta-regression in psychology, education, and medicine. *Research Synthesis Methods* **10**(2), 180–194.

Tsiatis, A. A. and Davidian, M. (2001) A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* **88**(2), 447–458.

Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G. and Higgins, J. P. T. (2012) Predicting the extent of heterogeneity in meta-analysis, using empirical data from the cochrane database of systematic reviews. *International Journal of Epidemiology* **41**(3), 818–827.

Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G. and Higgins, J. P. T. (2015) Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine* **34**(6), 984–998.

Van Houwelingen, H. C., Zwinderman, K. H. and Stijnen, T. (1993) A bivariate approach to meta-analysis. *Statistics in Medicine* **12**(24), 2273–2284.

Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P. T., Langan, D. and Salanti, G. (2016) Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods* **7**(1), 55–79.

Viechtbauer, W. (2005) Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics* **30**(3), 261–293.

Viechtbauer, W. (2007a) Confidence intervals for the amount of heterogeneity in meta-analysis. *Statistics in Medicine* **26**(1), 37–52.

Viechtbauer, W. (2007b) Hypothesis tests for population heterogeneity in meta-analysis. *British Journal of Mathematical and Statistical Psychology* **60**(1), 29–60.

Viechtbauer, W. (2010) Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software* **36**(3), 1–48.

Viechtbauer, W., López-López, J. A., Sánchez-Meca, J. and Marín-Martínez, F. (2015) A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological Methods* **20**(3), 360–374.

Walter, S. (1997) Variation in baseline risk as an explanation of heterogeneity in meta-analysis. *Statistics in Medicine* **16**(24), 2883–2900.

Wang, H., Boissel, J. P. and Nony, P. (2009) Revisiting the relationship between baseline risk and risk under treatment. *Emerging Themes in Epidemiology* **6**(1).

Whitehead, A. (2002) Dealing with heterogeneity. In *Meta-Analysis Of Controlled Clinical Trials*, ed. A. Whitehead. New Jersey: John Wiley & Sons, Ltd.

Xiong, J., Lipsitz, O., Nasri, F., Lui, L. M. W., Gill, H., Phan, L., Chen-Li, D., Iacobucci, M., Ho, R., Majeed, A. and McIntyre, R. S. (2020) Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders* **277**, 55–64.

Yi, G. Y. (2017) *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. First edition. New York: Springer.

Yi, G. Y., Delaigle, A. and Gustafson, P. (2021) *Handbook of Measurement Error Models*. First edition. Boca Raton: Chapman and Hall/CRC.

Thien Phuc Tran

CURRICULUM VITAE

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 371 335 2979
e-mail: thienphuc.tran@studenti.unipd.it

Current Position

Since October 2020; (expected completion: May 2024)
PhD Student in Statistical Sciences, University of Padova.
Thesis title: Advances in control risk regression...
Supervisor: Prof. Annamaria Guolo

Research interests

- Meta-analysis
- Measurement error

Education

Octorber 2018 – July 2020
Master (*laurea specialistica/magistrale***) degree in Mathematics** .
University of Verona, Faculty of computer science
Title of dissertation: "Improving the performance of classification using rank-based signature "
Supervisor: Prof. Mario Lauria
Co-supervisor: Prof. Luca Di Persio.
Final mark: 109

September 2014 – November 2014
Bachelor degree (*laurea triennale***) in Mathematics**.
Ho Chi Minh University of Science, Faculty of Mathematics
Title of dissertation: "Applying Machine Learning to Consumer Lending Analytics "
Supervisor: Dr. Duong Dang Xuan Thanh
Final mark: 9.04.

Visiting periods

February 2022 – July 2022
Brown School of Public Health,
Providence, US.
Supervisor: Prof. Christopher H. Schmid

Further education

None

Work experience

None

Awards and Scholarship

None

Computer skills

- Matlab
- R
- C++

Language skills

Vietnamese: native; English: good; Language3: fluent/good/moderate/basic (written/spoken).

Publications

Articles in journals
None

Chapters in books
None

Working papers
None

Conference presentations

Tran, P. T., Guolo, A. (2022). Likelihood-based inference in control risk regression with study-specific covariates. (poster) *CMStatistics Conference*, London, UK, 17-19th December.

Tran, P. T., Guolo, A. (2022). Likelihood-based inference in control risk regression with study-specific covariates. (poster) *CMStatistics Conference*, Newcastle, UK, 21-25 August.

Teaching experience

September 2022 – January 2023
Statistics for Management
Master's degree
Exercises, 30 hours
University of Padova
Instructor: Prof. Anna Giraldo

September 2021 – January 2022
Statistics for Management
Master's degree
Exercises, 30 hours
University of Padova
Instructor: Prof. Anna Giraldo

Other Interests

None

References

Prof. Annamaria Guolo
University of Padova
Address: 241 Cesare Battisti street, 35121 Padova city PD, Italy
Phone: ...
e-mail: annamaria.guolo@unipd.it

Prof. Christopher H. Schmid
Brown School of Public Health
Address: 121 South Main street, Providence, RI 02903, United States
Phone: ...
e-mail: christopher_schmid@brown.edu