Synthesis of regression results under different covariate sets

Daisuke Yoneoka

Doctor of Philosophy

Department of Statistical Science
School of Multidisciplinary Sciences
SOKENDAI (The Graduate University for Advanced Studies)

Synthesis of regression results under different covariate sets

by

Daisuke Yoneoka

yoneoka.daisuke@ism.ac.jp

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the
School of Multidisciplinary Sciences
Department of Statistical Science
SOKENDAI (The Graduate University for Advanced Studies)

March 2016

Abstract

Introduction:

Recently, increased development of clinical prediction models has been reported in the medical literature. However, evidence synthesis methodologies for these prediction models have not been sufficiently studied, especially for practical situations such as a meta-analyses where only aggregated summaries of important predictors are available. Also, in general, the covariate sets involved in the prediction models are not common across studies. As in ordinary model misspecification problems, dropping relevant covariates would raise potentially serious biases to the prediction models, and consequently to the synthesized results.

Methods:

We developed synthesizing methods for clinical prediction models with possibly different sets of covariates. In order to aggregate the regression coefficient estimates from different prediction models, we adopted a generalized least squares approach with non-linear terms (a sort of generalization of multivariate meta-analysis). Firstly, we evaluated omitted variable biases in this approach. Then, under an assumption of homogeneity of studies, we developed bias-corrected estimating procedures for regression coefficients of the synthesized prediction models.

Results:

Numerical evaluations with simulations showed that our approach resulted in smaller biases and more precise estimates compared with conventional methods, which use only studies with common covariates or which utilize a mean imputation method for omitted coefficients. These methods were also applied to several read dataset such as a series of Japanese epidemiologic studies on the incidence of a stroke.

Discussion:

Our proposed methods adequately correct the biases due to different sets of covariates between studies, and would provide precise estimates compared with

the conventional approach. If the assumption of homogeneity within studies is plausible, this methodology would be useful for incorporating prior published information into the construction of new prediction models.

Acknowledgements

First of all, I wish to express my great gratitude to my supervisor, Dr. Masayuki Henmi. He supervised all through my research during my PhD course. I could not finish this work without his valuable and sincere advices, considerable encouragement and endless support. Furthermore, I am grateful to the member of the dissertation committee, Professor Shinto Eguchi, Dr. Hisashi Noma and Dr. Kunihiko Takahashi for giving me much helpful advices and judging my PhD degree. In particular, Dr. Hisashi Noma gave me thoughtful and carefully structured comments and encouraged me to the last as my second supervisor. I also appreciate significant assistance and encouragement from the staff in the Institute of Statistical Mathematics and SOKENDAI.

Finally, I would express my special thanks to my colleagues in SOKENDAI and my family, Kazuya and Chiaki Yoneoka for their understanding and sincere and continuous support.

Contents

List of Figures

List of Tables

Abbreviations

EBM	**E**vidence **B**ased **M**edicine
RCT	**R**ondomized **C**ontrolled **T**rial
IPD	**I**ndividual **P**atient **D**ata
GLS	**G**eneralized **L**east **S**quares
GNLS	**G**eneralized **N**onlinear **L**east **S**quares
GLM	**G**eneralized **L**inear **M**odel
ML	**M**aximum **L**ikelihood
REML	**RE**stricted **M**aximum **L**ikelihood
MLE	**M**aximum **L**ikelihood **E**stimatior
JPHC study	**J**apan **P**ublic **H**ealth **C**enter-based **P**rospective **study**
DGP	**D**ata **G**enerating **P**rocess

This thesis is dedicated to my family and friends.

For their endless and thoughtful support and encouragement

Chapter 1

Introduction

The Cochrane collaboration [1] defines "meta-analysis" as follows:

> The use of statistical techniques in a systematic review to integrate the results of included studies. Sometimes misused as a synonym for systematic reviews, where the review includes a meta-analysis

The level of evidence of a meta-analysis is the highest level (rank 1a) in the field of evidence based medicine (EBM) [2].

Throughout the past several decades, meta-analysis has been widely accepted in many fields, including medicine, as a high-quality research methodology that quantitatively synthesizes the results of studies reporting on the same topic. In this chapter, I briefly outline and review the important meta-analysis methods, focusing on medical research.

1.1 Brief history

Gene Glass of the American Research Association [3] coined the term meta-analysis in 1976, to describe his basic quantitative method for synthesizing results. Meanwhile, Robert Rosenthal, John Hunter, and Frank Schmidt were

working on the same model for synthesizing study results [4]. Although meta-analysis was first discussed in the educational fields, its application has extended beyond education, and especially into medicine and health. As EBM has become increasingly relevant in the past 30 years, the demand and impact of meta-analysis has increased, especially in the medical fields. The number of meta-analysis publications in PubMed, provided by Suttton and Higgins [5], has increased by seven times from 1990 to 2006. Among these, the number of statistical papers published in the journal Statistics in Medicine has increased by five times (approximately 10 papers within 2012). This trend is continuing or even strengthening today [5, 4]. The Cochrane Collaboration endeavors to synthesize high-quality global evidence (for example, the results from randomized controlled trials (RCTs)) published in the Cochrane Database of Systematic Reviews. The Cochrane Library is an online collection of databases concerning the effectiveness of healthcare treatments and interventions. It contains over 2000 protocols and 5000 reviews on specific medical decisions.

1.2 What is meta-analysis from a statistical perspective?

In a meta-analysis, the results in each study $i, (i = 1, 2, \ldots, K)$ are evaluated by an effect size $\hat{\theta}_i$ such as standardized mean difference, odds ratio or p-value, which provides a common metric for each study [6, 7]. In addition, $\hat{\theta}_i$ is assumed to be normally distributed, which is reasonable when the effect size is the standardized mean difference transformed by the variance stabilizing procedure (as proposed by Hedges and Olkin [6]), but is unsuitable for log-odds ratios and crude correlation coefficients [6, 8].

The true effect size from the collection of studies can be estimated by the two proposed statistical models; the fixed effects model or the random effects model. The fixed effects model supposes that the studies are homogenous; therefore, the effect sizes of each study are presumed as unknown constants to be estimated [9, 8]. Conversely, in the random effects model, the effect sizes from individual

studies are randomly sampled from a distribution, and the hyperparameters in the distribution are estimated [8, 10].

1.2.1 Statistical inference in fixed-effects meta-analysis

When setting up a meta-analysis for survival data, we may reasonably assume that the log hazard ratio $\hat{\theta}_i$ and its variance $\hat{\sigma}_i^2$ are observable in study i ($i = 1, \ldots, K$). In a pioneering study, Hedges and Olkin inferenced the true effect size θ by calculating the weighted average of the effect sizes of individual studies, where the weight was the inverse of the variance [6].

$$\hat{\theta}_{fix} = \frac{\sum_{i=1}^{K} w_i \hat{\theta}_i}{\sum_{i=1}^{K} w_i}, \tag{1.1}$$

where $\hat{\theta}_{fix}$ is a fixed-effect estimator and the weights are the reciprocals of the reported variance:

$$w_i = \frac{1}{\hat{\sigma}_i^2}. \tag{1.2}$$

The variance of $\hat{\theta}_{fix}$ is calculated as

$$\mathrm{Var}(\hat{\theta}_{fix}) = \frac{1}{\sum_{i=1}^{K} w_i} \tag{1.3}$$

Because the $\hat{\theta}_i$ are normally distributed, it follows that $\hat{\theta}_{fix}$ is also normally distributed with mean θ and variance computed by Equation 1.3.

Intuitively, large-sample studies tend to have smaller variance than smaller studies, and are therefore weighted more heavily. In this way, the weights defined in Equation 1.2 control the variability of the estimated true effect size (i.e., variance of $\hat{\theta}_{fix}$).

Note that this model is acceptable when the studies are reasonably homogeneous. Homogeneity can be statistically estimated by the following test, or evaluated from other source information.

1.2.2 Statistical inference in random-effects meta-analysis

The fixed effects model supposes that the effect size $\hat{\theta}_i$ is fixed and unknown constant. Under this assumption, the variance of $\hat{\theta}_i$ is a single parameter, $\hat{\sigma}_i^2$. In the random effects model, the effect size $\hat{\theta}_i$ is assumed as random with a certain distribution, and $\hat{\theta}_i$ is commonly written as

$$\hat{\theta}_i = \theta_i + \epsilon_i = \theta + \xi_i + \varepsilon_i, \tag{1.4}$$

that is, the reported effect sizes are decomposed into fixed and random components. θ is the true effect, ξ_i is the random effect with mean 0 and unknown variance τ^2, and ε_i is the error term with mean 0 and known variance $\hat{\sigma}_i^2$. In this article, we adopt the commonly used terms within-study variance and between-study variance for the variances $\hat{\sigma}_i^2$ and τ^2, respectively.

In the other formulation of the random effects model, $\hat{\theta}_i$ is drawn from a distribution with mean θ_i and $\hat{\sigma}_i^2$; furthermore, each study-specific mean θ_i is assumed to follow a distribution with mean θ and variance τ^2:

$$\hat{\theta}_i|\theta_i, \hat{\sigma}_i^2 \sim N(\theta_i, \hat{\sigma}_i^2)$$
$$\theta_i|\theta, \tau^2 \sim N(\theta, \tau^2),$$

where θ and τ^2 are generally called hyperparameters. The posterior distribution of θ_i, conditional on the data and the hyperparameters θ and τ^2, is denoted as

$$\theta_i|\hat{\theta}_1, \ldots, \hat{\theta}_K, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2, \theta, \tau^2 \sim N(B_i\theta + (1 - B_i)\hat{\theta}_i, \hat{\sigma}_i^2(1 - B_i)), \tag{1.5}$$

where $B_i = \hat{\sigma}_i^2/(\hat{\sigma}_i^2 + \tau^2)$ is referred to as the shrinkage factor of the ith study [11].

Assuming that ξ_i and ε_i follow normal distributions, Equation 1.4 can be rewritten as

$$\hat{\theta}_i \sim N(\theta, (\hat{\sigma}_i^2 + \tau^2)^{-1}). \tag{1.6}$$

DerSimonian and Laird proposed an estimate that combines the reported effect sizes under this setting [10]; namely,

$$\hat{\theta}_{ran} = \frac{\sum_{i=1}^{K} w_i^* \hat{\theta}_i}{\sum_{i=1}^{K} w_i^*}, \tag{1.7}$$

where

$$w_i^* = \frac{1}{\hat{\sigma}_i^2 + \tau^2} \tag{1.8}$$

and the variance is given by

$$\mathrm{Var}(\hat{\theta}_{ran}) = \frac{1}{\sum_{i=1}^{K} w_i^*} \tag{1.9}$$

Here, the main interest is to inference the key parameter τ^2 in the random effects model. This is achieved by one of the four major methods: 1) the maximum likelihood (ML) method, 2) the restricted maximum likelihood (REML) method, 3) full Bayesian method, and 4) the moment estimation method [12, 11].

Note that by comparing the maximum log-likelihoods of this model and the fixed effects model described in the previous section, we can derive the likelihood ratio test for homogeneity of studies. This statistical test is discussed in the following section.

1.2.2.1 Maximum likelihood method

The log likelihood function of Equation 1.6 is given (up to a constant term) by

$$l(\theta, \tau^2)_{ML} = -\frac{1}{2} \sum_{i=1}^{K} \left[\log(\hat{\sigma}_i^2 + \tau^2) + \frac{(\hat{\theta}_i - \theta)^2}{\hat{\sigma}_i^2 + \tau^2} \right]. \tag{1.10}$$

In this subsection, the suffix ML represents the maximum likelihood method. For a balanced model, $\hat{\sigma}_i = \hat{\sigma}_1 = \text{const}$, the log likelihood (Equation 1.10) can be maximized in closed form with the following solutions:

$$\hat{\theta}_{ML} = \frac{1}{K} \sum_{i=1}^{K} \hat{\theta}_i, \quad \hat{\tau}^2_{ML} = \frac{1}{K} \sum_{i=1}^{K} (\hat{\theta}_i - \frac{1}{K} \sum_{i=1}^{K} \hat{\theta}_i)^2 - \hat{\sigma}_1^2. \tag{1.11}$$

If τ^2 is known, the maximum $l(\theta, \tau^2)_{ML}$ is expressed in the form of Equation 1.7. If $\hat{\sigma}_i^2 \neq$ const, the solutions must be inferred by iterative methods such as the Fisher scoring algorithm or the Newton-Raphson method. To apply an iterative inference method, we require the first and second derivatives; that is

$$\begin{aligned}
\frac{\partial l}{\partial \theta} &= \sum_{i=1}^{K} \frac{\hat{\theta}_i - \theta}{\hat{\sigma}_i^2 + \tau^2}, \quad \frac{\partial l}{\partial \tau^2} = \sum_{i=1}^{K} \left[\frac{1}{\hat{\sigma}_i^2 + \tau^2} - \frac{(\hat{\theta}_i - \theta)^2}{(\hat{\sigma}_i^2 + \tau^2)^2} \right] \\
\frac{\partial^2 l}{\partial \theta^2} &= -\sum_{i=1}^{K} \frac{1}{\hat{\sigma}_i^2 + \tau^2}, \quad \frac{\partial^2 l}{\partial \tau^4} = \frac{1}{2} \sum_{i=1}^{K} \left[\frac{1}{(\hat{\sigma}_i^2 + \tau^2)^2} - \frac{2(\hat{\theta}_i - \theta)^2}{(\hat{\sigma}_i^2 + \tau^2)^3} \right] \\
\frac{\partial^2 l}{\partial \theta \tau^2} &= -\sum_{i=1}^{K} \frac{\hat{\theta}_i - \theta}{(\hat{\sigma}_i^2 + \tau^2)^2}.
\end{aligned}$$

Therefore, the Hessian matrix, $\boldsymbol{H}$, for l is denoted as

$$\boldsymbol{H} = -\begin{pmatrix} \sum_{i=1}^{K} \frac{1}{\hat{\sigma}_i^2 + \tau^2} & \sum_{i=1}^{K} \frac{\hat{\theta}_i - \theta}{(\hat{\sigma}_i^2 + \tau^2)^2} \\ \sum_{i=1}^{K} \frac{\hat{\theta}_i - \theta}{(\hat{\sigma}_i^2 + \tau^2)^2} & -\frac{1}{2} \sum_{i=1}^{K} \left[\frac{1}{(\hat{\sigma}_i^2 + \tau^2)^2} - \frac{2(\hat{\theta}_i - \theta)^2}{(\hat{\sigma}_i^2 + \tau^2)^3} \right] \end{pmatrix}. \quad (1.12)$$

Because the $(2, 2)$ element in the matrix $\boldsymbol{H}$ may be positive, the matrix is not negative semi definite. Therefore, the log likelihood function Equation 1.10 is not a concave function, and the Newton-Raphson algorithm can fail by converging to a local maximum.

The negative expected Hessian matrix, called the information matrix, $\mathcal{I}$, is given by

$$\mathcal{I} = -E(\boldsymbol{H}) = \begin{pmatrix} \sum_{i=1}^{K} \frac{1}{\hat{\sigma}_i^2 + \tau^2} & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^{K} \frac{1}{(\hat{\sigma}_i^2 + \tau^2)^2} \end{pmatrix}. \quad (1.13)$$

Unlike the Hessian matrix $\boldsymbol{H}$, the information matrix $\mathcal{I}$ is always positive definite. Consequently, the Fisher scoring algorithm is more stable than the Newton-Raphson algorithm [12].

Regarding the maximum likelihood estimators (MLEs) of θ and τ^2, the expected cross-derivative is 0, so the MLEs of θ and τ^2 are asymptotically independent for large numbers of studies (i.e., large K). Therefore, the log likelihood

functions of θ and τ^2 can be separately maximized in practice. Note that in the meta-analysis setting, the asymptotic theory should generally be based on the number of studies K, not the number of samples in each study.

Inverting the information matrix, $\mathcal{I}$, yields the asymptotic covariance matrix. The lower Cramer-Rao bounds for the asymptotic estimates of θ and τ^2 are then written as

$$\mathrm{Var}(\hat{\theta}^2_{ML}) = \left(\sum_{i=1}^{K} \frac{1}{(\hat{\sigma}_i^2 + \tau^2)} \right)^{-1} \tag{1.14}$$

$$\mathrm{Var}(\hat{\tau}^2_{ML}) = 2 \left(\sum_{i=1}^{K} \frac{1}{(\hat{\sigma}_i^2 + \tau^2)^2} \right)^{-1}. \tag{1.15}$$

1.2.2.2 Restricted maximum likelihood method

As it is well known, the maximum likelihood estimation of the variance is biased under finite sampling. Therefore, Laird and Ware proposed the REML method for linear mixed models [13]. Under the meta-analysis setting of random effects, the maximum likelihood function contains an additional term, $-\frac{1}{2}\log\sum(\hat{\sigma}_i^2 + \tau^2)^{-1}$, as

$$l(\theta, \tau^2)_{REML} = -\frac{1}{2} \sum_{i=1}^{K} \left[\log(\hat{\sigma}_i^2 + \tau^2) + \frac{(\hat{\theta}_i - \theta)^2}{\hat{\sigma}_i^2 + \tau^2} + \log \sum_{i=1}^{K} \frac{1}{\hat{\sigma}_i^2 + \tau^2} \right]. \tag{1.16}$$

In this subsection, the suffix, $REML$, represents the restricted maximum likelihood method. In a balanced model, $\hat{\sigma}_i^2 = \hat{\sigma}_1^2 = \text{const}$, $\hat{\theta}_{REML} = K^{-1}\sum_{i=1}^{K}\hat{\theta}_i$ and

$$\hat{\tau}^2_{REML} = \frac{1}{K-1} \sum_{i=1}^{K} (\hat{\theta}_i - \frac{1}{K} \sum_{i=1}^{K} \hat{\theta}_i)^2 - \hat{\sigma}_1^2. \tag{1.17}$$

Note that although the REML estimator of the variance is unbiased in the balanced model (unlike the ML estimator Equation 1.11), the bias still remains in an unbalanced model. On the other hand, the estimator of the true (population)

mean (i.e., hyperparameter θ) in the unbalanced model can be calculated as

$$\hat{\theta}_{REML} = \frac{\sum_{i=1}^{K} w_{REML,i}\hat{\theta}_i}{\sum_{i=1}^{K} w_{REML,i}}, \tag{1.18}$$

where $w_{REML,i} = \dfrac{1}{\hat{\sigma}_i^2 + \hat{\tau}^2_{REML}}$. The estimator of θ_i is obtained by plugging the REML estimator into Equation 1.5, yielding an approximation known as empirical Bayes. However, further consideration is needed because the empirical Bayes approximation fixes the hyperparameters θ and τ^2, ignoring their uncertainties [11].

1.2.2.3 Full Bayesian method

The full Bayesian method incorporates the uncertainty in the hyperparameters θ and τ^2 [11, 14].

The inference of the true (population) mean θ (and the mean from each study θ_i) can be calculated by integrating out the unknown parameters. For this purpose, we define $\theta \sim N(0, a^2)$ and $\tau^{-2} \sim gamma(c, d)$, and denote the joint posterior distribution of $\theta, \theta_1, \ldots, \theta_K, \tau^2$ as [11];

$$p(\theta, \theta_1, \ldots, \theta_K, \tau^2 | \hat{\theta}_1, \ldots, \hat{\theta}_K, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2) \propto \prod_{i=1}^{K} p(\theta_i | \hat{\theta}_i, \hat{\sigma}_i^2) p(\theta_i | \theta, \tau^2) p(\theta) p(\tau^2). \tag{1.19}$$

The inferences are made from this posterior distribution as

$$\begin{aligned} \hat{\theta}_{FB} &= E(\theta | \hat{\theta}_1, \ldots, \hat{\theta}_K, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2) \\ &= \int_{\theta} \int_{\theta_i} \int_{\tau^2} \theta \; p(\theta, \theta_1, \ldots, \theta_K, \tau^2 | \hat{\theta}_1, \ldots, \hat{\theta}_K, \hat{\sigma}_1^2, \ldots, \hat{\sigma}_K^2) \mathrm{d}\theta_i \mathrm{d}\tau^2 \mathrm{d}\theta. \end{aligned} \tag{1.20}$$

The suffix, FB, represents the full Bayesian method. Typically, this calculation cannot be written in closed form unless the prior distribution and the likelihood are conjugate, but the posterior distribution can be inferred by Monte Carlo method or some other methods [10, 15].

1.2.2.4 Moment estimation method

To derive an estimator of τ^2, DerSimonian and Laird [10] applied a moment estimation method using the statistics from a homogeneity test (Equation 1.29, detailed in the next section), which is given by

$$\tau^2_{MM} = \max\left\{0,\ \frac{Q-(K-1)}{\sum_{i=1}^{K} w_i - \frac{\sum_{i=1}^{K} w_i^2}{\sum_{i=1}^{K} w_i}}\right\}, \tag{1.21}$$

where $w_i = \dfrac{1}{\hat{\sigma}_i^2}$. The suffix, MM means the moment estimation method.

The estimator was then obtained as follows:

$$\hat{\theta}_{MM} = \frac{\sum_{i=1}^{K} w_{MM,i}\hat{\theta}_i}{\sum_{i=1}^{K} w_{MM,i}}, \tag{1.22}$$

where $w_{MM,i} = \dfrac{1}{\hat{\sigma}_i^2 + \hat{\tau}^2_{MM}}$.

1.2.3 Multivariate random effect meta-analysis

Recently, multivariate meta-analysis with fixed or random effects has received much attention, despite being more complex than univariate meta-analysis. Multivariate meta-analysis can jointly analyze multiple and correlated outcomes. The clinical applications of multivariate meta-analysis are detailed in Riley et al. [16]. Here I focus on the random effects multivariate meta-analysis model because of its popularity among practitioners.

For simplicity, let us consider K studies ($i = 1, \ldots, K$), each with two outcomes of interest. In the ith study, denoted by $\hat{\theta}_{ij}$ and $\hat{\sigma}^2_{ij}$ the jth effect size ($j = 1, 2$) and its associated variance, respectively. Both summary statistics, $\hat{\theta}_{ij}$ and $\hat{\sigma}^2_{ij}$, are assumed to be known and $\hat{\theta}_{ij}$ is an estimate of the true effect size θ_{ij}. In addition, θ_{ij} is assumed to independently follow a certain distribution with overall (or mean) effect size θ^*_j and between-study variance τ^2_j (where both $\hat{\theta}_{ij}$ and θ_{ij} follow normal distributions) [17].

Under these settings, the random effects meta-analysis model becomes

$$\hat{\boldsymbol{\theta}}_i = \begin{pmatrix} \hat{\theta}_{i1} \\ \hat{\theta}_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_{i1}^2 & \hat{\sigma}_{i1}\hat{\sigma}_{i2}\rho_{W,i} \\ \hat{\sigma}_{i1}\hat{\sigma}_{i2}\rho_{W,i} & \hat{\sigma}_{i2}^2 \end{pmatrix} \right) \tag{1.23}$$

$$\boldsymbol{\theta}_i = \begin{pmatrix} \theta_{i1} \\ \theta_{i2} \end{pmatrix} \sim N\left(\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \tau_1\tau_2\rho_B \\ \tau_1\tau_2\rho_B & \tau_2^2 \end{pmatrix} \right), \tag{1.24}$$

where Equation 1.23 and Equation 1.24 represent the within-study and between-study structures, respectively, and ρ_{Wi} and ρ_B are the respective within-study and between-study correlations. Thus, by aggregating Equation 1.23 and Equation 1.24, we can use the marginal distribution of $\hat{\theta}_{i1}$ and $\hat{\theta}_{i2}$ for the inference on θ_j and ρ_j, $j = 1, 2$ (note that $\hat{\sigma}_{ij}^2$ is assumed as a known constant), that is

$$\begin{pmatrix} \hat{\theta}_{i1} \\ \hat{\theta}_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \boldsymbol{V}_i = \begin{pmatrix} \hat{\sigma}_{i1}^2 + \tau_1^2 & \hat{\sigma}_{i1}\hat{\sigma}_{i2}\rho_{W,i} + \tau_1\tau_2\rho_B \\ \hat{\sigma}_{i1}\hat{\sigma}_{i2}\rho_{W,i} + \tau_1\tau_2\rho_B & \hat{\sigma}_{i2}^2 + \tau_2^2 \end{pmatrix} \right) \tag{1.25}$$

The log restricted likelihood of $\theta_1, \theta_2, \tau_1^2, \tau_2^2, \rho_B$ is given by

$$\begin{aligned} & l(\theta_1, \theta_2, \tau_1^2, \tau_2^2, \rho_B) \\ & = -\frac{1}{2}\left[\log(|\sum_{i=1}^{K} \boldsymbol{V}_i^{-1}|) + \sum_{i=1}^{K} \left\{ \log|\boldsymbol{V}_i| + (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta})^T \boldsymbol{V}_i^{-1} (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}) \right\} \right] \end{aligned} \tag{1.26}$$

This likelihood can be maximized by the REML method as described in Van Houwelingen et al. and others [18, 12, 16]. However, the REML method has two major drawbacks [19, 16, 17, 20]: 1) knowledge about ρ_B [19, 20] is rarely available and 2) the estimated covariance matrix may be singular, leading to biased estimates of the standard errors and confidence intervals [17, 20].

1.2.3.1 Riley's method

Riley et al. [20] proposed a new method that overcomes the above-mentioned problems. This method requires no knowledge of the correlations among the

outcomes ρ_{Wi}, and probably avoids the convergence problem when the covariance matrix is singular.

Instead of using τ_i^2, ρ_{Wi}, and ρ_B, Riley et al. [20] introduced a new correlation parameter of synthesis ρ_S, which accounts for the marginal correlation between $\hat{\theta}_{i1}$ and $\hat{\theta}_{i2}$. The method is given by

$$\hat{\boldsymbol{\theta}}_i = \begin{pmatrix} \hat{\theta}_{i1} \\ \hat{\theta}_{i2} \end{pmatrix} \sim N\left(\boldsymbol{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \boldsymbol{\Psi}_i = \begin{pmatrix} \hat{\sigma}_{i1}^2 + \psi_1^2 & \rho_S\sqrt{(\psi_1^2 + \hat{\sigma}_{i1}^2)(\psi_2^2 + \hat{\sigma}_{i2}^2)} \\ \rho_S\sqrt{(\psi_1^2 + \hat{\sigma}_{i1}^2)(\psi_2^2 + \hat{\sigma}_{i2}^2)} & \hat{\sigma}_{i2}^2 + \psi_2^2 \end{pmatrix}\right), \tag{1.27}$$

where ψ_j^2 represents a new variation added to the within-study variation and ρ_S is a correlation parameter for the marginal distribution (in general, ρ_S and the between-study variance τ_j^2 are not equivalent). To infer $\theta_1, \theta_2, \psi_1^2, \psi_2^2$, and ρ_S, they applied the restricted maximum likelihood method to the log likelihood, defined as

$$\begin{aligned} &l(\theta_1, \theta_2, \psi_1^2, \psi_2^2, \rho_S) \\ &= -\frac{1}{2}\left[\log(|\sum_{i=1}^{K} \boldsymbol{\Psi}_i^{-1}|) + \sum_{i=1}^{K}\left\{\log|\boldsymbol{\Psi}_i| + (\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta})^T \boldsymbol{\Psi}_i^{-1}(\hat{\boldsymbol{\theta}}_i - \boldsymbol{\theta})\right\}\right]. \end{aligned} \tag{1.28}$$

1.2.4 Quantifying heterogeneity

The studies included in meta-analysis generally differ in their designs and implementations as well as in their sample sizes, interventions and outcomes. Such methodological heterogeneity leads to value discrepancies in the reported results of collections of studies, a condition called heterogeneity [5, 21]. Heterogeneity in meta-analysis complicates the synthesis and its implication, impairing researchers' ability to draw overall conclusions. If significant variability is detected, researchers should seek potential moderate variables and try to explain this variability.

The consistency of the results from each study can be visualized by useful tools: a forest plot or some other graphical method [21, 22, 23]. L'abbe plots

can also help to detect unusual studies when each reported result is formatted as a 2×2 table [24]. When the homogeneity assumption seems plausible (or fails to reject the null hypothesis of homogeneity in the following statistical test for heterogeneity), researchers usually apply a fixed-effects model because the differences in the estimated effect sizes among the studies arise only from sampling error. On the other hand, when the homogeneity assumption seems unrealistic, a random effects model is more appropriate [25].

Traditionally, statistically significant heterogeneity among studies has been determined by Cochran's Q statistic [26, 27]. As a chi-squared test, the Q test is easily computed by summing the squared deviations of the estimates, weighting the contribution of each study by its inverse variance.

$$Q = \sum_{i=1}^{K} w_i (\hat{\theta}_i - \hat{\theta}_{meta})^2, \tag{1.29}$$

where the weights w_i are defined in Equation 1.2 and $\hat{\theta}_{meta}$ means the estimates from meta-analysis by using methods explained above. Under the null hypothesis of effect-size homogeneity, the Q statistics follow a chi-squared distribution with $K - 1$ degrees of freedom (where K is the number of studies). Thus, if the Q statistic exceeds the critical value for a given significance level α (ex. $= 0.05$ or 0.01), the null hypothesis is rejected and the heterogeneity among the studies is concluded as statistically significant. However, several studies have reported flaws in the Q statistic. Theoretical proofs have established that because its statistical power depends on the number of studies, the Q statistic performs poorly at detecting true heterogeneity when few studies are included in the meta-analysis, and detects low heterogeneity among many studies [8, 28]. Second, the Q test only informs us of heterogeneity among included studies, but does not report the extent of such heterogeneity [25].

Heterogeneity among the studies in a meta-analysis can also be evaluated by the H^2, R^2 and I^2 measures, proposed by Higgins and Thompson [29]. Here, I focus on I^2 because this measure is popular among practitioners. The I^2 index quantifies the degree of heterogeneity in included studies by dividing the difference between the actual and expected Q values by the actual Q value (where

the expected Q value is the number of degrees of freedom $K-1$) [25]:

$$I^2 = \begin{cases} \dfrac{Q-(K-1)}{Q} * 100 & \text{if } Q \geq K-1 \\ 0 & \text{if } Q < K-1. \end{cases} \tag{1.30}$$

This index is easily interpreted as the proportion of the total variation that can be explained by the between-studies variance:

$$I^2 = \frac{\hat{\tau}^2}{\hat{\tau}^2 + \hat{\sigma}^2}. \tag{1.31}$$

In other words, the I^2 index is the percentage of the heterogeneity (between-study variance) in the total variability. Higgins and Thompson proposed a classification criterion for I^2 values, by which practitioners can easily interpret the magnitude of the heterogeneity (25%, 50%, and 75% represent low, medium, and high heterogeneity, respectively) [29]. They also theoretically derived the confidence intervals by alternatively calculating the H^2 index as $H^2 = \dfrac{Q}{K-1}$. The H^2 index and I^2 indexes are closely related through

$$I^2 = \frac{H^2-1}{H^2}. \tag{1.32}$$

Therefore, we can calculate the confidence intervals of H^2 instead of those of I^2. Higgins and Thompson [29] assumed that the natural logarithm of H has a standard normal distribution; thus the confidence intervals are given by

$$\exp\left\{\log(H) \pm \|z_{\alpha/2}\| SE(\log(H))\right\}, \tag{1.33}$$

where $\|z_{\alpha/2}\|$ is the quantile of the standard normal distribution (which depends on the significance level α) and SE means the standard errors. They also estimated

$$SE(\log(H)) = \begin{cases} \dfrac{1}{2}\dfrac{\log(Q)-\log(K-1)}{\sqrt{2Q}-\sqrt{2K-3}} & \text{if } Q > K \\ \sqrt{\dfrac{1}{2(K-2)}\left(1-\dfrac{1}{3(K-2)^2}\right)} & \text{if } Q \leq K. \end{cases} \tag{1.34}$$

The confidence interval of I^2 is obtained by combining Equation 1.32, Equation 1.33 and Equation 1.34.

However, although the I^2 index is widely recognized as an alternative Q statistic, it performs with low power when the number of studies is small, as demonstrated in Monte Carlo simulations [25].

1.3 Meta-analysis of regression results

Regression analyses, particularly multiple regressions, are fundamental statistical tools for understanding the associations between a variable of interest and its outcome. For example, Elmore and Woehlke [30] showed that, from 1978 to 1987, approximately 20% of the articles published in several education journals (Educational Researcher, American Educational Research Journal, and Review of Educational Research) used a regression approach to associate the wages, educational histories and quality of teachers and their test scores. This trend has continued in recent years. Another example is the meta-analysis of clinical prediction models in medicine. Clinical prediction models have been increasingly used to assess various diseases. Systematic reviews have reported 102 risk prediction models for cardiovascular disease [31] and 25 models for detecting the risk factors of type 2 diabetes (including 11 logistic regression models) [32].

Almost all sophisticated research designs and statistical analyses estimate the true effect size of interest by controlling or partitioning out the effects of other variables. However, methodologies for synthesizing complex regression models have not kept pace with the exploding supply of results and the increasing demand for combined research [33]. In practice, most meta-analyses ignore such studies and instead summarize the bivariate relationships, measured by indices such as correlations and standardized mean differences. Consequently, many of the articles reporting recent advances and efforts have been largely ignored.

1.3.1 Methods for synthesizing regression results

Now I review some proposed methods for combining regression results. Especially, the method of Becker and Wu has motivated our extended synthesis method [33].

In this subsection, I presume that K studies are involved in the meta-analysis. For each study i, I define an estimated coefficients vector $\hat{\boldsymbol{\beta}}_i, (i = 1, \dots K)$ and its covariance matrix $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_i)$.

1.3.1.1 Summaries of t statistics

Walker and Saw [34] and Stanley and Jarel [35] newly proposed the t statistics for synthesizing regression results. The t statistic, defined as the slope divided by its standard error, is provided in most popular statistical packages. Presumably, this metric avoids the problem of different scales of variables across candidate studies. However, this method is not without problems [33]. Most seriously, the t statistic includes not only the magnitude of the effect size, but also the sample size and model precision. Therefore, the t statistic is vulnerable to changes in the estimates and their standard errors, both of which easily occur under changes of the sample size and/or residual variation of the regression.

When evaluating the impact of political advertisements, Lau et al. [36] synthesized the t statistics to express the mean difference between groups. The rationale for their approach was tackling the large discrepancy among the candidate models. They found that 1/4 of studies were based on ordinary least squares methods or logistic regression. By using the t statistic, Lau et al. could accommodate the different models across the studies. The t statistic is also employed in Timm's index; the so-called ubiquitous effect size [37]. Given $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)^T$, an $n \times p$ design matrix $\boldsymbol{X} = (\mathbf{1}, \boldsymbol{X}_2, \dots, \boldsymbol{X}_{p-1})$ and a $p \times 1$ parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})$, we construct a linear model for $\boldsymbol{Y}$ as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$, where $\boldsymbol{e} \sim N(0, \sigma^2)$. Suppose that we are interested in the effect size of the coefficient β_1. In this case, we can write $\boldsymbol{C}\boldsymbol{\beta} = (0\ 1\ 0 \dots\ 0)\boldsymbol{\beta} = \beta_1$. The Timm's index incorporates this vector function into the hypothesis test

$H_0 : \boldsymbol{\Psi} = \boldsymbol{C\beta} = \varphi_0$. The Timm's index is written as

$$\Delta_T = \frac{\left\{(\boldsymbol{\Psi} - \varphi_0)^T (\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T)^{-1}(\boldsymbol{\Psi} - \varphi_0)\right\}^{\frac{1}{2}}}{\sigma\sqrt{(n-p)+q+1}}, \tag{1.35}$$

where $\boldsymbol{C}$ is a $q \times p$ matrix consisting of linear combinations of β_0 to β_{p-1}, and q is the rank of $\boldsymbol{C}$. This index relaxes the dependency of the t statistics on the sample size by incorporating a multiplier that mitigates the influence of the sample size. However, Timm provided no detailed method for combining the index.

1.3.1.2 Synthesis of dose response models

Greenland [38], Greenland and Longnecker [39], and Shi and Copas [40] conducted a meta-analysis of the regression coefficients in dose-response models. Mainly, they were interested in the positive associations between risks and exposure levels (for example, between the number of cigarettes smoked per day and the risk of lung cancer). The results of such studies are typically synthesized by random effects modeling weighted by the within-study variance. The regression models in various studies tend to have a common covariate (dose level), so the regression coefficients are easily synthesized by multivariate meta-analysis. However, the exposure levels of the subjects are frequently grouped into categories (such as high, middle, and low dose groups) or intervals rather than individually recorded, which confounds the meta-analysis.

To tackle this problem, Shi and Copas [40] treated dose as a continuous variable with observable intervals. They proposed a maximum likelihood method for estimating the mean dose response relationship and the between-study variance structure of the regression coefficients. They also proposed a homogeneity test for the dose response curve based on likelihood ratios. They assumed that if there exists a single covariate (a variable indicating the exposure level), every available dose response model is a bivariate regression model. They argued that their method approximates the adjusted odds ratio if the influence of the other adjusted covariates is sufficiently small. However, they did not present an exact methodology, which would synthesize such regression models without any approximations.

1.3.1.3 Synthesis with IPD

Multivariate meta-analysis can synthesize the correlated effects from multiple outcomes. Such joint synthesis improves the efficiency of the meta-analysis, because the multivariate meta-analysis can borrow strength from other various correlated outcomes [41]. Multivariate meta-analysis is frequently used to synthesize regression coefficients because if the fitted models in several studies yield the same regression outcomes, the coefficients of those models can be correlated and the each regression result can be used to borrow its strength [41]. Unfortunately, the within-study correlation [19, 42, 20, 43], which is required to fit the multivariate meta-analysis model, is rarely reported. However, the within-study correlation can be calculated from the IPD of each study (if available).

Multivariate meta-analysis based on IPD (IPD meta-analysis) is implemented in two frameworks: 1) the familiar two-stage estimation framework [44, 45], and 2) one-stage estimation [46, 47], a new inference method that constructs more exact likelihood functions, but which is limited to special cases.

I first explain the two-stage estimation method. In the first stage, each study is separately analyzed by its IPD; in the second stage, the aggregated first-stage results are subjected to a standard multivariate meta-analysis. The first stage of IPD meta-analysis fits each regression result to a common model with a covariate set available in every IPD. The fitted model should be carefully chosen by variable selection methods and model building. After fitting the common model to each IPD, the regression results of study i (estimated coefficients $\hat{\boldsymbol{\beta}}_i (i = 1, \dots K)$ and their covariance matrix $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_i)$) are stored for the second-stage analysis.

At the second stage of IPD meta-analysis, the multivariate meta-analysis introduced in 1.2.2 is implemented. For example, assume that two covariates are commonly available in each IPD, and that each IPD can be fitted to a common linear model $Y_i = \alpha_{0i} + \alpha_{1i} X_{i1} + \alpha_{2i} X_{i2} + e_i$. The coefficients and covariate distributions are assumed to differ among the models (indicated by the subscript i in the coefficients). In these settings, the second-stage IPD meta-analysis of

the regression coefficients (in the ordinary case) is given by

$$\hat{\boldsymbol{\beta}}_i = \begin{pmatrix} \hat{\beta}_{i0} \\ \hat{\beta}_{i1} \\ \hat{\beta}_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_{i0} \\ \beta_{i1} \\ \beta_{i2} \end{pmatrix}, \mathrm{Cov}(\hat{\boldsymbol{\beta}}_i) \right) \tag{1.36}$$

$$\boldsymbol{\beta}_i = \begin{pmatrix} \hat{\beta}_{i0} \\ \hat{\beta}_{i1} \\ \hat{\beta}_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_0\tau_1\rho_{B01} & \tau_0\tau_2\rho_{B02} \\ \tau_0\tau_1\rho_{B01} & \tau_1^2 & \tau_2\tau_3\rho_{B12} \\ \tau_0\tau_2\rho_{B02} & \tau_2\tau_3\rho_{B12} & \tau_2^2 \end{pmatrix} \right). \tag{1.37}$$

The ML or REML method easily infers the parameters of interest such as the mean (global average) $\beta_0, \beta_1, \beta_2$ and structure $\tau_0, \tau_1, \tau_2, \rho_{B01}, \rho_{B02}, \rho_{B12}$ of a random effect. However, although the random effect model can incorporate study heterogeneity, it obscures the objective of a meta-analysis of regression results because it computes the global average effect, which may not be identical across studies. Thus, what the average effect means in a practical setting requires careful consideration [40]. Moreover, an IPD meta-analysis easily obtains the within-study correlation (i.e., $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_i)$), but in most cases, the IPD is unavailable and the covariance matrix of coefficients is not reported. Therefore, it is necessary to impute or infer the missing elements of the unreported covariance matrix of coefficients from the available information. The missing elements of the covariance matrix are frequently the off-diagonal elements (i.e., the correlations among the coefficients).

The second framework is the one-stage method for IPD meta-analysis, which does not require the within-study correlations. Assuming that the outcomes are mutually exclusive or have "is subset of" relationship, Trikalions et al. [47] proved that the number of events in a single study can be modeled using multinomial distributions. They represented the heterogeneity among studies by random effects, and constructed the likelihood. They noted several advantages of this one-stage estimation method; 1) it is available in several statistical software packages [46, 48], 2) the multinomial distribution models the exact structure, avoiding the need for a large-sample approximation to the normal likelihood (as in standard multivariate meta-analysis), and 3) it requires no prior knowledge of the within-study correlation. However, this method is limited to meta-analyses of

count data, and is inapplicable to regression coefficients because it assumes an underlying multinomial distribution of the number of events [47]. Therefore, this method is inappropriate for my present study.

1.3.1.4 Generalized least squares approach

Becker and Wu proposed a synthesis method based on generalized least squares (GLS), which was first outlined by Raudenbush et al. for calculating standardized mean differences [33, 49].

First, the K sets of coefficients reported in meta-analysis studies are stacked $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1^T, \ldots, \hat{\boldsymbol{\beta}}_K^T)^T$ and a blockwise diagonal matrix is constructed from the reported covariance matrices of coefficients $\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Cov}(\hat{\boldsymbol{\beta}}_1) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \mathrm{Cov}(\hat{\boldsymbol{\beta}}_K) \end{pmatrix}$.

Then, assuming that the K studies have a common set of P covariates and that each slope vector $\hat{\boldsymbol{\beta}}_i$ estimates the true parameter $\boldsymbol{\beta}$, the model is expressed as

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_{10} \\ \hat{\beta}_{11} \\ \vdots \\ \hat{\beta}_{1P} \\ \vdots \\ \hat{\beta}_{KP} \end{pmatrix} = \boldsymbol{W}\boldsymbol{\beta} + \boldsymbol{e} = \left(\begin{array}{cccc} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 \\ \hline \vdots & \vdots & \vdots & \vdots \\ \hline 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 \end{array}\right) * \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_P \end{pmatrix} + \boldsymbol{e}, \tag{1.38}$$

where the design matrix $\boldsymbol{W}$ is a K-times piled identity matrix composed of 0s and 1s, and $\boldsymbol{e}$ follows a normal distribution with mean zero and variance $\boldsymbol{\Sigma}$. By the GLS technique, $\boldsymbol{\beta}$ and its covariance are estimated as

$$\hat{\boldsymbol{\beta}}^* = (\boldsymbol{W}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{W})^{-1}\boldsymbol{W}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\beta}} \tag{1.39}$$

and

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}^*) = (\boldsymbol{W}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{W})^{-1}, \tag{1.40}$$

where $\hat{\boldsymbol{\beta}}^*$ represents the synthesized estimator.

The GLS approach seems theoretically appealing because it handles the unequal variances of effects in different-sized studies. In addition, by recognizing the error term as the within-study variance, this approach formulates a multivariate meta-analysis with fixed effects. Thus, we can acquire new insight using GLS-based statistical tools, which have accumulated discussions on heteroscedasticity and robust methods.

1.4 Motivation studies

In this section, I introduce two studies whose approaches have motivated my study.

First, my study extends the work of Becker and Wu [33], which was mentioned in section 1.3.1.4. Their GLS-based approach to synthesizing regression results is a variant of multivariate meta-analysis with fixed effect. However, this approach is limited in practice for two reasons: 1) in practical situations, the covariate sets should vary among studies and 2) the covariance matrices of coefficients are supposed to be reported, which is rarely the case. Regarding the first problem (i.e., different sets of covariates), if the covariates in models are imbalanced, their associated coefficients are difficult to synthesize because their interpretation depends on which covariates are included in each regression model. Previous attempts to solve this problem are discussed in the next section. The second limitation of Becker and Wu's study (unreported covariance matrices of coefficients) is similar to the problem of unreported within-study correlations, which is frequently discussed in researches of multivariate meta-analysis. Standard multivariate meta-analysis assumes that the within-study correlations are known or can be deduced from the IPDs. Previous studies relevant to this issue are discussed in section 2.5. In addition, although there are several methods for recovering or inserting elements in the unknown covariance matrix of coefficients,

there are no directly applicable techniques for recovering the exact values of unreported covariate correlations from the summary statistics; all such methods require the IPD. The present study extends Becker and Wu's method [33] by recovering the correlation estimates. The recovering procedure is presented in this thesis.

Second, my work was motivated by Debray et al.'s study [50]. These authors newly conceptualized the synthesis of prediction models, and evaluated several approaches to aggregating previously published prediction models with a new IPD. Several different approaches were applied to 15 datasets of traumatic brain injury, and to prediction models of deep venous thrombosis. They concluded that synthesizing the prediction models improves the discrimination ability and calibration of the final models under various scenarios. However, they did not consider the imbalance of covariates in each model; that is, they gave identical interpretations of the regression coefficients in different models, even when the covariate sets differed among the models. In this case, the differences in the covariance sets must be incorporated as an additional term in the synthesis model. That is, the interpretation of the coefficients depends on the covariates included in the model.

1.5 Statistical methodologies for missing covariates in meta-analysis

In the previous section, I mentioned that sets of covariates may vary across candidate models. As in the ordinary misspecification problem, ignoring this differences biases the synthesized results because the missing covariates might be correlated. In this study, the bias is regarded as an omitted variable bias. For example, let us assume 6 studies with different covariate sets having a nested and monotone missing structure (i.e., studies 1 and 2 analyze the full model, and the remaining studies analyze covariate subsets of the full model). As mentioned elsewhere, if the full models (studies 1 and 2) represent the true model, the other models are misspecified and their coefficients will differ from the true parameters

of interest in the full model (i.e., $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4$).

$$\textbf{Study 1} \quad Y_1 = \alpha_{01} + \alpha_{11}X_{11} + \alpha_{21}X_{21} + \alpha_{31}X_{31} + \alpha_{41}X_{41} + e_1$$
$$\textbf{Study 2} \quad Y_2 = \alpha_{02} + \alpha_{12}X_{12} + \alpha_{22}X_{22} + \alpha_{32}X_{32} + \alpha_{42}X_{42} + e_2$$
$$\textbf{Study 3} \quad Y_3 = \beta_{03} + \beta_{13}X_{13} + \beta_{23}X_{23} + \beta_{33}X_{31} + e_3$$
$$\textbf{Study 4} \quad Y_4 = \gamma_{04} + \gamma_{14}X_{14} + \gamma_{24}X_{24} + e_4$$
$$\textbf{Study 5} \quad Y_5 = \gamma_{05} + \gamma_{15}X_{15} + \gamma_{25}X_{25} + e_5$$
$$\textbf{Study 6} \quad Y_6 = \delta_{06} + \delta_{16}X_{16} + e_5$$

In this example, most researchers would consider the full model (studies 1 and 2) as the most powerful and informative model; thus, the full model can reasonably be regarded as the true model.

As a more practical example, consider the following non-nested structure of the covariate sets:

$$\textbf{Study 1} \quad Y_1 = \alpha_{01} + \alpha_{11}X_{11} + \alpha_{21}X_{21} + \alpha_{31}X_{31} + \alpha_{41}X_{41} + e_1$$
$$\textbf{Study 2} \quad Y_2 = \alpha_{02} + \alpha_{12}X_{12} + \alpha_{22}X_{22} + \alpha_{32}X_{32} + \alpha_{42}X_{42} + e_2$$
$$\textbf{Study 3} \quad Y_3 = \beta_{03} + \beta_{23}X_{23} + \beta_{33}X_{33} + \beta_{43}X_{41} + \beta_{53}X_{51} + e_3$$
$$\textbf{Study 4} \quad Y_4 = \gamma_{04} + \gamma_{24}X_{24} + \gamma_{44}X_{44} + \gamma_{64}X_{64} + e_4$$
$$\textbf{Study 5} \quad Y_5 = \gamma_{05} + \gamma_{25}X_{25} + \gamma_{45}X_{45} + \gamma_{64}X_{64} + e_4$$
$$\textbf{Study 6} \quad Y_6 = \delta_{06} + \delta_{16}X_{16} + e_5$$

In both cases, it is necessary to determine the true model, and aggregate the regression results into that model. Therefore, I strongly recommend determining the most appropriate model as true model in advance. The rest of this section focuses on meta-analyses of regression results with missing variables, as indicated in the above examples.

A naive approach to this problem is a multivariate meta-analysis on studies with common sets of covariates, excluding all other studies. In the Simulation section, the naive method provides a comparison group for checking the performance of the proposed method. Theoretically, the naive approach introduces

no bias to the synthesized results, but its efficiency is degraded by ignoring the indirect information from studies with different types of covariate sets.

The second simple approach was proposed by Debray et al. [50]. They adopted a mean or zero imputation method with large variance (= 100) for missing coefficients. They compared this approach, which they called uninformative regression coefficients, with the above naive method on studies with full sets of covariates. However, unless the omitted covariates have little influence on the remaining covariates or outcomes, this imputation method biases the synthesized results. Note that a package **mvmeta** in both R and STATA include functions treating this missing issue; *inputna* function replaces missing covariances with 0, and missing variances with the largest observed variance. Under the assumption of missing at random (MAR), Wu and Becker proposed the factored likelihood method to calculate synthesized standardized slope by using sweep operator [51]. They put relatively strong assumption that their are only monotone missing structure and that data rom each study are standardized with mean zero and one standard deviation for each variables. But under the settings of fixed-effect model and monotone missing structure, their method can be viewed as variant of the zero imputation method which is frequently used in R and STATA like above [51].

Here I present a linear regression, but when the meta-analysis synthesizes models with nonlinear or GLM formulations, the omitted covariates should exert a nonlinear influence [52, 53, 54, 55]. In this study, such cases are treated as special examples of a logistic regression model.

The other approach are based on the use of IPD. Fibrinogen Studies Collaboration (FSC) proposed a multivariate meta-analysis approach to borrow strength from partially adjusted results based on IPD [56]. This approach was demonstrated in practice by Riley et al. [44]. For simplicity, the FSC considers two types of Cox regression models, given by

$$\text{Full proportional hazard model} \quad \lambda(t) = \lambda^f \exp(\beta_1^f X_1 + \beta_2 X_2)$$
$$\text{Partial proportional hazard model} \quad \lambda(t) = \lambda^p \exp(\beta_1^p X_1),$$

where λ is the hazard function and t indicates time [57]. In the full model, the superscript f indicates quantities that are fully adjusted for X_1 and X_2, whereas the superscript p indicates quantities that are partially adjusted without the covariates X_2. This notation emphasizes the differences in the coefficients of X_1 (i.e., β_1) and in the baseline hazard model. Both models are assumed to be fitted to a cohort and to permit estimation of the coefficients. Lastly, assume that we seek the influence of the covariate X_1; that is, β_1 is the parameter of interest.

For a given cohort, the FSC method assumes that

$$\begin{pmatrix} \hat{\beta}_1^f \\ \hat{\beta}_1^p \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_1^f \\ \beta_1^p \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right), \tag{1.41}$$

where σ_1^2, σ_2^2, and ρ are known parameters. This situation mirrors the usual univariate meta-analysis, in which the within-study variance is assumed fixed and known. However, as ρ is rarely reported, estimating its inference presents an inherent difficulty. The FSC method estimates ρ by a Taylor's expansion, as described below.

Second, the FSC method assumes that β_1^f and β_1^p have underlying distributions (exerting a random effect on the parameters). Therefore, these parameters should vary among cohorts as

$$\begin{pmatrix} \beta_1^f \\ \beta_1^p \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_f \\ \beta_p \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \kappa\tau_1\tau_2 \\ \kappa\tau_1\tau_2 & \tau_2^2 \end{pmatrix} \right). \tag{1.42}$$

After calculating the marginal distribution by Equation 1.41 and Equation 1.42, we obtain the following standard bivariate meta-analysis model:

$$\begin{pmatrix} \hat{\beta}_1^f \\ \hat{\beta}_1^p \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_f \\ \beta_p \end{pmatrix}, \begin{pmatrix} \sigma_1^2 + \tau_1^2 & \rho\sigma_1\sigma_2 + \kappa\tau_1\tau_2 \\ \rho\sigma_1\sigma_2 + \kappa\tau_1\tau_2 & \sigma_2^2 + \tau_2^2 \end{pmatrix} \right), \tag{1.43}$$

The number of models K includes the numbers of full ($i = 1, \ldots, k^*$) and partial ($i = k^* + 1 \ldots, K$) models. Under these settings, the log likelihood function is

obtained as

$$L(\beta_f, \beta_p, \tau_1^2, \tau_2^2, \kappa) = \sum_{i=1}^{k^*} \log f_i(\hat{\beta}_1^f, \hat{\beta}_1^p) + \sum_{i=k^*+1}^{K} \log f_i(\hat{\beta}_1^p), \tag{1.44}$$

where $f_i(\hat{\beta}_1^f, \hat{\beta}_1^p)$ and $f_i(\hat{\beta}_1^p)$ are obtained from Equation 1.43. This calculation is easily implemented in standard meta-analysis techniques such as REML.

The FSC study also proposed two inference methods for estimating the unknown correlations in each cohort; 1) nonparametric bootstrapping and 2) Taylor expansion. The first is the simplest and most intuitive approach, but is slow. The bootstrap sample is taken from participants in the IPD, which provides the information for X_2 with replacement [58]. At each bootstrap sample, the paired estimates $\hat{\beta}_1^{f*}$ and $\hat{\beta}_1^{p*}$ are stored and used to inference the bootstrap estimates ρ_b. Inference by Taylor's expansion is relatively analytical and yields an approximation only. This procedure is essentially a variant of Steyerberg et al.'s method [59, 60]. The FSC method assumes a linear relationship between $\hat{\beta}_1^f$ and $\hat{\beta}_1^p$; namely, $\hat{\beta}_1^p = \hat{\beta}_1^f + \hat{\beta}_2\hat{\gamma}$, where $\hat{\gamma}$ are the estimated regression coefficients of X_2 on X_1. This linear relationship indicates that

$$\mathrm{Cov}(\hat{\beta}_1^f, \hat{\beta}_1^p) = \mathrm{Cov}(\hat{\beta}_1^f, \hat{\beta}_1^f + \hat{\beta}_2\hat{\gamma}) = \mathrm{Var}(\hat{\beta}_1^f) + \mathrm{Cov}(\hat{\beta}_1^f, \hat{\beta}_2\hat{\gamma}).$$

As the $Y|X_1, X_2$ and $X_2|X_2$ are independent, ($\hat{\beta}_1^f$ and $\hat{\beta}_2$ are functions of $Y|X_1, X_2$ and $\hat{\gamma}$ is a function of $X_2|X_1$), $\mathrm{Cov}(\hat{\beta}_1^f, \hat{\beta}_2\hat{\gamma}) = \mathrm{Cov}(\hat{\beta}_1^f, \hat{\beta}_2 E[\hat{\gamma}])$. Using the approximation $E[\hat{\gamma}] \approx \hat{\gamma}$, the covariance of the coefficient between the full and partial model is approximated as

$$\mathrm{Cov}(\hat{\beta}_1^f, \hat{\beta}_1^p) \approx \mathrm{Var}(\hat{\beta}_1^f) + \mathrm{Cov}(\hat{\beta}_1^f, \hat{\beta}_2)\hat{\gamma},$$

where the term on the right-hand-side is obtained from the cohort with full covariates sets (i.e., the information of the cohort with covariate X_2). The within-study correlation ρ_a is calculated as

$$\rho_a = \frac{\mathrm{Cov}(\hat{\beta}_1^f, \hat{\beta}_1^p)}{\left\{\mathrm{Var}(\hat{\beta}_1^f)\mathrm{Var}(\hat{\beta}_1^p)\right\}^{\frac{1}{2}}}.$$

As mentioned above, the FSC developed IPD-based synthesis of regression results. Additionally, Resche-Rigon et al. [61] adopted a multiple imputation method with IPD to correct the omitted variable bias (in studies of missing imputation, this omission is called systematic missing). Multiple imputation is an attractive approach for handling missing data [62, 63, 64]; thus, it might also be applicable to meta-analyses of different covariate sets across studies. Several multiple imputation methods have been proposed for sporadically missing covariates; that is, when the covariates are incompletely observed in each cluster (referred to as partially observed covariates in meta-analysis) [61]. The first and simplest approach is to separately impute the missing covariates within each study. Another approach is to impute the missing covariates among all studies within a single scheme. However, the hierarchical data structure (multilevel data) in the meta-analysis setting must be carefully preserved in the imputation procedure. To resolve this difficulty, Schafer and Yucel proposed a Gibbs sampler for multiple imputation using multivariate linear mixed models [65]. Van Buuren introduced the multiple imputation by chained equation (MICE) technique for such multilevel data, which is based on the variable-by-variable method [66]. They employed linear random effect models and developed a Gibbs sampler of the posterior distribution of the missing data.

To cope with systematically missing covariates (i.e., varying covariate sets across studies), Resche-Rigon et al. [61] employed multilevel random effects models in an extended MICE approach. They considered the following Cox model with random effects on the intercept and slope:

$$\lambda_i(t) = \lambda_0(t) \exp(\sum_{r=1}^{k} \beta_r x_{ir} + \sum_{r=1}^{l} u_r x_{ir} + u_0), \tag{1.45}$$

where $\lambda_i(t)$ denotes the hazard function at time t in study $i = 1, \ldots, K$. Some covariates are assumed as missing. The existing covariates are ordered under the rule that their coefficients $l \leq k$ have underlying distributions (i.e., random effects). The fixed coefficients are denoted β_r, and u_r and u_0 are random effects with $u \sim MVN(0, \tau)$ (where MVN denotes the multivariate normal distribution). The rth covariate is assumed to be systematically missing the continuous variable x_{ir}. Resche-Rigon et al. presumed heterogeneity in both the variable means and

the associations between variables. Therefore, they rationalized that the linear mixed-effect model was an appropriate imputation model. This mixed model allows random intercepts and random effects on covariates; thereby, the imputed values can be estimated from the distribution inferred from other studies.

Let x_{ir} be a $n_i \times 1$ vector in study i. The following linear relationship is assumed:

$$x_{ir} = W_{ir}\gamma_{ir} + Z_{ir}b_{ir} + e_{ir}, \tag{1.46}$$

where W_{ir} is an $n_i \times p$ matrix representing the fixed effect of p covariates. Included in W_{ir} are other covariates $x_{is}(s \neq r)$ and the outcome functions. γ_{ir} is a $p \times 1$ vector of fixed-effect parameters, Z_{ir} is a $n_i \times q$ matrix representing the random effect, and $b_{ir} \sim N(0, \Psi_r)$ is a $q \times 1$ vector of random effects (Ψ_r is a $q \times 1$ covariance matrix embodying the normal distribution of the random effect), $e_{ir} \sim N(0, \sigma_r I)$ is a $n_i \times 1$ vector of residuals, and σ_r is the residual variance parameter. Under these settings and assuming that data are MAR, proper imputation is performed in the following steps:

1. Calculate the MLEs of the parameters $\hat{\gamma}_r, \hat{\Psi}_r, \hat{\sigma}_r$ in the imputation model, using the information from studies including the covariate x_{ir}
2. Sample γ_r^* from the distribution $N(\hat{\gamma}_r, \mathrm{Var}(\hat{\gamma}_r|\hat{\Psi}_r, \hat{\sigma}_r))$
3. Sample Ψ_r^*, σ_r^* from the distribution $N((\hat{\Psi}_r, \hat{\sigma}_r), \mathrm{Var}(\hat{\Psi}_r, \hat{\sigma}_r))$
4. Sample b_{ir}^* from the distribution $N(0, \Psi_r^*)$ of study i, which is systematically missing covariate x_{ir}
5. Sample x_{ir}^* from the distribution $e \sim N(0, \sigma_r^*)$ and construct the equation

$$x_{ir}^* = W_{ir}\gamma_{ir}^* + Z_{ir}b_{ir}^* + e_{ir}^*$$

   for study i with systematically missing covariate x_{ir}.

The MAR assumption means that whether or not the x_{ir} are systematically missing from a study is independent of x_{ir} itself, and conditional on the observed

data in that study. More precisely, whether an x_{ir} is systematically missing is not determined from the mean of that variable or its associations with other variables in the study. For selecting W_{ir}, the matrix should include all covariates and outcomes [63, 67]. Especially, the outcome of the analysis model (the target model of the research) should be included in the imputation model [68]. The selection of Z_{ir} is more complicated. Resche-Rigon et al. [61] proposed the following 3 models for inferring Z_{ir}:

- Model 1: No random effect model (i.e., $Z_{ir} = 0$)
- Model 2: A random intercept model (i.e., Z_{ir} includes only one constant term)
- Model 3: Model 2 plus a random effect on some of the covariates

1.6 Statistical methodology for misspecified estimating functions

This subsection focuses on misspecified models and their influence on an estimating function (for details, see [53]). Grace and Reid [53] extended the work of White [55], who related misspecified models and maximum likelihood estimates to Kullback-Leibler divergence. In this thesis, the regression coefficient synthesis is based on the following theorems (1.6.1 and 1.6.2) and the techniques of [53].

Suppose that I wish to estimate a vector θ of interest from a sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ ($i = 1, \ldots, n$), drawn from the density $\{f(\boldsymbol{y}; \theta)\}$. The dimensions of $\boldsymbol{y}_i$ and θ are d and p, respectively. Given a $p \times 1$ vector of estimating functions $g(\boldsymbol{y}; \theta)$, $\hat{\theta}_n$ is the root of the following equation:

$$G_n(\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{y}_i; \hat{\theta}) = \mathbf{0}. \tag{1.47}$$

Provided that the model conditions are regular and the estimating functions are unbiased; that is, $E_\theta[g(\boldsymbol{y}_i); \theta] = \mathbf{0}$, $\hat{\theta}_n$ is consistent and asymptotically normal

with an asymptotic covariance matrix. Such a matrix is known as the Godambe information matrix as [69];

$$J^{-1}(g) = \left\{ E_\theta \left[\frac{\partial g}{\partial \theta^T} \right] \right\}^{-1} E_{gg^T}[gg^T] \left\{ E_\theta \left[\frac{\partial g^T}{\partial \theta} \right] \right\}^{-1}. \tag{1.48}$$

Biased estimating functions (which are quite common) are denoted as $h(\boldsymbol{y}; \theta)$ (i.e., $E_\theta[h(\boldsymbol{y}_i); \theta] \neq 0$). The bias in $h(\boldsymbol{y}; \theta)$ is most simply derived by modifying the estimating function. To this end, $E_\theta[h(\boldsymbol{y}_i; \theta)]$ is calculated as

$$\hat{H}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} h(\boldsymbol{y}_i; \theta) - E_\theta[h(\boldsymbol{y}_i; \theta)].$$

The calculation of $E_\theta[h(\boldsymbol{y}_i; \theta)]$ poses the greatest difficulty. This term may be approximated by methods such as bootstrapping [70] and the moment method [71]. These methods are based on a theoretical result; namely that the unbiasedness of the estimating function relates to the conditional likelihood inference in exponential families [71].

A more explicit inference method is illustrated in [53]. Instead, of Equation 1.47, let us assume that

$$H_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} h(\boldsymbol{y}_i; \theta) = \mathbf{0}, \tag{1.49}$$

has a root $\hat{\theta}_n^*$ for any sample $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, and that

$$E_\theta[h(\boldsymbol{y}; \theta^*)] = \mathbf{0} \tag{1.50}$$

is satisfied for some θ^*. The root of Equation 1.50, $\hat{\theta}_n^*$ is a function of the true parameter θ (i.e., $\hat{\theta}_n^* = \hat{k}(\theta)$). If the inverse function

$$\theta = k(\theta^*) \tag{1.51}$$

exists, a new estimator of θ can be constructed from Equation 1.51 as

$$\hat{\hat{\theta}}_n = k(\hat{\theta}_n^*). \tag{1.52}$$

Next, I briefly explain the asymptotic results of this estimator; that is, the consistency and asymptotic normality. Detailed proofs are provided in [53]. The first theorem proves the convergence of the estimator $\hat{\theta}_n^*$, which is derived from estimating functions that may be biased. According to Grace and Reid [53],

Theorem 1.6.1.

Suppose $h(\boldsymbol{y};\theta) = \{h_1(\boldsymbol{y};\theta), \ldots, h_p(\boldsymbol{y};\theta)\}^T$ *is a vector of functions such that* $h_j(\boldsymbol{y};\theta)(j = 1, \ldots, p)$ *is a continuous function of* θ *for each* $\boldsymbol{y}$ *and a measurable function of* $\boldsymbol{y}$ *for each* θ. *Also suppose that* Θ *is a convex and compact set and that the true distribution of* $\boldsymbol{Y}$ *is* $F = F(\boldsymbol{y};\theta_0)$ *with density function* $f(\boldsymbol{y};\theta_0)$ *for some* $\theta_0 \in \Theta$. *Assume that* $m_j()$ *is an integrable function with respect to* F *and that* $|h_j(\boldsymbol{y};\theta)| \leq m_j(\boldsymbol{y})$ *for all* $\boldsymbol{y}$ *and* θ. *Let* $H(\theta) = E_{\theta_0}[h(\boldsymbol{Y};\theta)]$ *and* $H_n(\theta) = \frac{1}{n}\sum_{i=1}^n h(\boldsymbol{y}_i;\theta)$. *If* $H(\theta) = \mathbf{0}$ *has a unique solution* θ_0^* *and* $H_n(\theta) = 0$ *has a solution* $\hat{\theta}_n^*$, *then*

$$\hat{\theta}_n^* \xrightarrow{p} \theta_0^* \quad as \quad n \to \infty$$

for almost every sequence $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ *randomly sampled from* F.

This theorem confirms the consistency property. The difference $\theta_0^* - \theta_0$ can be considered as the asymptotic bias sourced from the biased estimating functions when estimating the true parameter θ. If the estimating function $h(\boldsymbol{Y};\theta)$ is unbiased, it is easily checked that $\theta_0^* = \theta_0$ and that $\hat{\theta}_n^*$ is consistent for θ.

The next theorem relates to the asymptotic normality property. Under the following settings

$$\begin{aligned}
A_n(\theta) &= \frac{1}{n}\sum_{i=1}^n \frac{\partial\{h(\boldsymbol{Y}_i;\theta)\}^T}{\partial\theta} \\
A(\theta) &= E_{\theta_0}[A_n(\theta)] \\
B_n(\theta) &= \frac{1}{n}\sum_{i=1}^n \{h(\boldsymbol{Y}_i;\theta)\}\{h(\boldsymbol{Y}_i;\theta)\}^T \\
B(\theta) &= E_{\theta_0}[B_n(\theta)] \\
C_n(\theta) &= A_n^{-1}(\theta)B_n(\theta)A_n^{-T}(\theta) \\
C(\theta) &= A^{-1}(\theta)B(\theta)A^{-T}(\theta),
\end{aligned}$$

we can derive the following theorem (for a detailed proof, see [53]).

Theorem 1.6.2.

Assume that the conditions of the above theorem are satisfied and that $h_j(\boldsymbol{y};\theta)$ *is a continuously differentiable function with respect to* θ *for each* $\boldsymbol{y}$*. Further assuming that* $A(\theta_0^*)$ *is nonsingular (i.e.,* $A(\theta_0^*)$ *is invertible), and imposing some regularity conditions on* $h_j(\boldsymbol{y};\theta)$ *and the model* F*, the following results are obtained:*

1. $\sqrt{n}(\hat{\theta}_0^* - \theta_0^*) \xrightarrow{d} N(\mathbf{0}, C(\theta_0^*))$
2. $C_n(\hat{\theta}_n^*) \xrightarrow{p} C(\theta_0^*)$ *and assume that* k *(defined in Equation 1.51), exists and is differentiable with respect to* θ*,*

$$\sqrt{n}(\hat{\theta}_0 - \theta_0) \xrightarrow{d} N\left(\mathbf{0}, \left(\frac{\partial k(\theta_0^*)^T}{\partial \theta}\right) C(\theta_0^*) \left(\frac{\partial k(\theta_0^*)}{\partial \theta^T}\right)\right)$$

Based on this theorem, we can construct the confidence intervals and hypothetical test. The regulatory conditions in both of the above theorems are similar to those of Van der Waart [72] (i.e., the first and second moments of h and $\frac{\partial h}{\partial \theta^T}$ exist, the estimating functions are differentiable with respect to θ, and the expectation and deviation are exchangeable). The compactness assumption of Θ can be relaxed [53] as similarly described in Huber [73] and Walker [74].

1.7 Problems and Aims

Methods for summarizing simple indices such as correlations, proportions, and mean differences have been studied and developed over the past 30 years. When single indices adequately represent the study outcomes, the results are easily synthesized by statistical packages. However, when combining the results of regression models, especially multivariate regression models (which are generally used to associate covariates with target outcomes of interest), methodologies for synthesizing the results are insufficient and the results are not well understood.

Considering increasing needs for statistical methodology of meta-analysis of clinical prediction models, the development of such meta-analysis methodologies is urgently needed. However, the regression settings present several difficulties. In practice, each study commonly has various sets of covariates, because researchers want to construct regression models on their own datasets with their original covariates, and this leads to the case where the covariate sets will differ among researches. Such variation causes differences in the estimable parameters and their interpretations. Debray et al. [50] alleviated this problem by specifying mean or zero imputation for the missing coefficients. However, their approach biases the synthesized coefficients because the interpretation of the coefficients depends on the covariates included in the models. On the other hand, conventional meta-analysis generally excludes such studies from the synthesis, and performs multivariate meta-analysis using only those studies with common covariates structures. However, such exclusion loses indirect information, degrading the efficiency of the multivariate meta-analysis.

To resolve these difficulties, this study proposes a new meta-analysis method for synthesizing regression coefficients. The method is designed for different sets of covariates in a more general setting. The coefficients are synthesized by a generalized non-linear least square (GNLS) method with bias correction terms. The different sets of covariates among studies are considered as analogous to the omitted variable bias (or the model misspecification problem). In this context, the statistical handling of misspecified estimating functions (section 1.6) is a suitable approach.

Chapter 2

Methods

In this chapter, firstly I describe the issue of omitted variable bias as a model misspecification problem. Conceptually, omitted variable bias can be regarded as an analogy of the issue of different covariates set in meta-analysis. Then, I propose a new method for synthesis of coefficients using the derived omitted variable bias formula. Lastly, I provide brief explanation about the properties of our estimator such as a robustness and an asymptotic normality. To check our methodology works, I provide several concrete examples such as a logistic model and linear model.

2.1 Settings

I consider a similar situation as Debray et al. [50]; I can use reported summary statistics from previous regression models with different sets of covariates and at least one IPD from the publications or the authors themselves. Suppose that each published prediction model has a subset of covariates in the IPD, and is constructed for same prediction task. The number of published prediction models is K $(i = 1, \ldots, K)$ and the ith article reports the estimated coefficients, $\hat{\boldsymbol{\theta}}_i$, and the covariance matrix $\boldsymbol{\Sigma}_i = \mathrm{Cov}(\hat{\boldsymbol{\theta}}_i)$ (at least its diagonal elements). Each $\hat{\boldsymbol{\theta}}_i$ is a column vector of possibly different length, $\hat{\boldsymbol{\theta}}_i = (\hat{\theta}_{i1}, \ldots, \hat{\theta}_{ip_i})^T$, and p_i represents the differences of covariate sets among studies.

To synthesize these regression coefficients, Debray et al. [50] utilize the mean or zero imputation methods for omitted coefficients and apply the technique of multivariate meta-analysis as described in the previous section. Another simple approach is to apply multivariate meta-analysis using studies with common covariate sets [50]. However, the former approach leads to biased results and the latter is not biased but leads to loss of efficiency by ignoring indirect information from omitted studies. In order to improve these methods, I propose a new method for synthesis of regression coefficients under different sets of covariates.

For simplicity, I assume the case where the true model has the full set of covariates in the IPD, which means the prior models have subsets of covariates and are considered as under-specified models. Note that since the omitted variables from the true models (full models) are correlated with the included variables, the subset models are confounded and biased compared with true models. Our method can be generalized to more complex cases where the previous models for meta-analysis are a mixture of under- and over-specified models.

2.2 Omitted variable bias in GLM

According to the original work of GLM by Nelder and Wedderburn [75], the outcome distribution is in an exponential family, such as the Gaussian, binomial, Poisson, gamma, or inverse-Gaussian distributions. Subsequently, GLM has been extended to multivariate exponential families (such as the multinomial distribution) and non-exponential families (such as the two-parameter negative-binomial distribution) [76]. This section will focus on the exponential family case for simplicity, which can be written in general expression as

$$f(y_i;\theta) = \exp\left(\frac{y_i\theta - \kappa(\theta)}{\phi} + c(y_i,\phi)\right), \tag{2.1}$$

where ϕ is the nuisance parameter (also called the dispersion parameter), θ is the parameter of interest (called the natural parameter), c is the normalization constant, and $\kappa(\theta)$ is the partition function. To convert from the mean parameter (i.e., the expectation $\mu = E[Y]$ of Y) to the natural parameter (or the canonical

parameter), we can use a link function $\Psi()$,

$$\theta = \Psi(\mu),$$

which function is uniquely determined by the form of the exponential family distribution. The term generalized linear model is used because covariates $\boldsymbol{X} = (X_1, \ldots, X_m)$ (independent variables) are linearly connected to the linear predictor η with the parameter of interest $\boldsymbol{w} = (w_1, \ldots, w_m)$ as

$$\eta = \boldsymbol{X}\boldsymbol{w},$$

which $\boldsymbol{w}$ is the alternative parameter of interest, and η is connected through the link function $\eta = g(\mu)$. When $g = \Psi$, it is called the canonical link function. For example, the logistic regression case sets $\mu_i = g^{-1}(\eta_i) = sigm(\eta_i)$, where $sigm$ represents the sigmoid function.

To derive the omitted variable bias, the unbiasedness condition of the estimating function can be employed [53] as briefly explained in the section 1.6, whose approach is a generalized result of the landmark paper of White [55]. In the meta-analysis framework, the idea of the omitted variable bias can work as an analogy of different covariate sets, and due to this property it also works as a representation of the incorporation of indirect information from prior models.

In the context of GLM, one of the candidate of estimating functions is the score function. Firstly, here I introduce the score function for GLM. For simplicity, I suppose the nuisance parameter ϕ is known and constant although many theories remains true without this restriction. Based on the results on Appendix B, I have

$$\begin{aligned} E[y_i|\boldsymbol{x}_i, w, \phi] &= \kappa^{'}(\theta) \\ \mathrm{Var}[y_i|\boldsymbol{x}_i, w, \phi] &= \kappa^{''}(\theta)\phi, \end{aligned}$$

where $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{mi})$ is a sample vector from the random vector $\boldsymbol{X}$.

The log-likelihood can be written as

$$L(\beta, \phi) = \sum_{i=1}^{N} \left\{ \frac{y_i \theta_i - \eta(\theta_i)}{\phi} + c(y_i, \phi) \right\} = \sum_{i=1}^{N} l_i, \tag{2.2}$$

where θ has a relationship to β through $g(\kappa^{'}(\theta_i)) = \eta_i = \boldsymbol{x}_i \boldsymbol{w}$.

Therefore, by the chain rule for differentiation and using the above, I have the score function as

$$\begin{aligned}
\frac{l_i}{w_j} &= \frac{dl_i}{d\eta_i} \frac{\partial \eta_i}{\partial w_i} \\
&= \frac{dl_i}{d\eta_i} \frac{\partial \eta_i}{\partial \theta_i} \frac{\partial \theta_i}{\partial w_j} \\
&= \frac{dl_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial w_j} \\
&= \frac{dl_i}{d\theta_i} \left(\frac{d\mu_i}{d\theta_i} \right)^{-1} \left(\frac{d\eta_i}{d\mu_i} \right)^{-1} \frac{\partial \eta_i}{\partial w_j} \\
&= \frac{y_i - \kappa^{'}(\theta_i)}{\phi} (\kappa^{''}(\theta_i))^{-1} (g^{'}(\mu_i))^{-1} x_{ij} \\
&= \frac{(y_i - \mu_i) x_{ij}}{\phi \mathrm{Var}(\mu_i) g^{'}(\mu_i)} \\
&= \frac{(y_i - g^{-1}(\boldsymbol{x}_i \boldsymbol{w})) x_{ij}}{\phi \mathrm{Var}(g^{-1}(\boldsymbol{x}_i \boldsymbol{w})) g^{'}(\mu_i)},
\end{aligned}$$

where $\mathrm{Var}(\mu_i) = \dfrac{d\mu_i}{d\theta_i}$ and this score function can be regarded as the estimating function, and x_{ij} indicates the value of covariate at the i'th individual and the j'th covariate.

Example 2.1. *Normal y_i with the mean $\mu_i = \boldsymbol{x}_i \boldsymbol{w}$*

In this case, $g(\mu_i) = \mu_i = \boldsymbol{x}_i \boldsymbol{w}$ holds, thus I have $g^{'}(\mu_i) = \dfrac{dg(\mu_i)}{d\mu_i} = 1$, $\mathrm{Var}(\mu_i) = 1$ and $\phi = \sigma^2$. Therefore, the log likelihood function can be reduced into

$$\frac{1}{\sigma^2} \sum_{i=1}^{N} (y_i - \boldsymbol{x}_i \boldsymbol{w}) x_{ij} = 0, \tag{2.3}$$

which can be written in the matrix form such as the j'th element of the vector $\boldsymbol{X}^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{w})$. Here, I define $\boldsymbol{Y}, \boldsymbol{X}$ as the outcome vector of y_i and the design

matrix of $\boldsymbol{x}_i$

Example 2.2. *Poisson* y_i *with the log link function*

In this case, $\mu_i = \log(\boldsymbol{x}_i\boldsymbol{w})$ *(i.e.,* $g(\mu_i) = \log(\mu_i)$*) holds, thus I have* $g^{'}(\mu_i) = \frac{dg(\mu_i)}{d\mu_i} = \frac{1}{\mu_i}$, $\mathrm{Var}(\mu_i) = \mathrm{Var}(y_i) = \mu_i$ *and* $\phi = \sigma^2$*. Therefore, the log likelihood function can be reduced into*

$$\sum_{i=1}^{N}(y_i - \exp(\boldsymbol{x}_i\boldsymbol{w}))x_{ij} = 0, \tag{2.4}$$

where there are not explicit solution for $\boldsymbol{w}$ *in general.*

In general, score functions from misspecified models cannot satisfy the unbiasedness condition of estimating functions (score functions). Suppose the data generating process (DGP) can be formulated by the true model

$$\eta = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{Z}\boldsymbol{\beta}, \tag{2.5}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the true parameters of interest, and also suppose the misspecified models is fitted without the covariate $\boldsymbol{Z} = (Z_1, \ldots, Z_l)$ as follow;

$$\eta = \boldsymbol{X}\boldsymbol{\alpha}. \tag{2.6}$$

The first and crucial step is to find the solution of the unbiasedness condition of estimating function (the score functions) derived from the misspecified model (Equation 2.6), i.e., find $\gamma^* = f(\alpha, \beta, p_{XZ})$, which is the function of the true parameters α, β and the joint distribution of covariates, p_{XZ};

$$E\left[\frac{(y_i - g^{-1}(\boldsymbol{x}_i\boldsymbol{\alpha}))x_{ij}}{\phi\mathrm{Var}(g^{-1}(\boldsymbol{x}\boldsymbol{\alpha}))g^{'}(\mu_i)}\right] = \boldsymbol{0}. \tag{2.7}$$

This expectation should be calculate with respect to the true distribution. The true distribution can be written as

$$p_{Y\boldsymbol{X}\boldsymbol{Z}} = \exp\left(\frac{Y\theta - \eta(\boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{Z}\boldsymbol{\beta})}{\phi} + c(Y, \phi)\right). \tag{2.8}$$

2.3 Nonlinear model for meta-analysis

Suppose there exist K reported models ($i = 1, \ldots, K$) with their estimated coefficients of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ and their covariance matrices, and when $i = 1, \ldots, k$, studies fit the true model (1) with a full set of covariates, $\boldsymbol{X}$ and $\boldsymbol{Z}$, and when $i = k+1, \ldots, K$, studies mistakenly omit covariates $\boldsymbol{Z}$. I assume the homogeneity of studies (i.e., the distribution of covariates and outcomes are common across the studies in the meta-analysis). Here I show only the case where $\boldsymbol{Z}$ is omitted, but the case where $\boldsymbol{X}$ is omitted can be considered in the same manner, and further, it is easy to generalize to various other omittance patterns. To synthesize the estimated coefficients vectors from the GLM models, I apply a GNLS method to incorporate the unequal variances of studies into meta-analysis.

Based on this setting, the nonlinear model for meta-analysis can be formulated as follows;

$$\hat{\boldsymbol{\theta}}_i = \boldsymbol{g}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, p_{\boldsymbol{XZ}}) + \boldsymbol{\varepsilon}_i \qquad (i = 1, \ldots, K), \tag{2.9}$$

where

$$\boldsymbol{g}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, p_{\boldsymbol{XZ}}) = \begin{cases} (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T & (i = 1, \ldots, k) \\ \boldsymbol{f}(\boldsymbol{\alpha}, \boldsymbol{\beta}, p_{\boldsymbol{XZ}}) & (i = k+1, \ldots, K), \end{cases},$$

$$\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_K \end{pmatrix} \sim N(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{Cov}(\hat{\boldsymbol{\theta}}_1) & \ldots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathrm{Cov}(\hat{\boldsymbol{\theta}}_K) \end{bmatrix},$$

and $\hat{\boldsymbol{\theta}}_i$ is the column vector of reported coefficients in the ith study. The function $\boldsymbol{f}()$ comes from the omitted variable bias formula introduced in the previous section, whose formulation is reasonable if an assumption of homogeneity of studies in meta-analysis is acceptable.

In a large sample, the estimated coefficients $\hat{\boldsymbol{\theta}}_i$ are (approximately) normally distributed with mean $\boldsymbol{\theta}_i = \boldsymbol{g}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, p_{\boldsymbol{XZ}})$ and covariance $\mathrm{Cov}(\hat{\boldsymbol{\theta}}_i)$. This asymptotic normality of estimated coefficients leads to the justification of the GNLS approach.

Under the model, overall estimates of the regression coefficients $\hat{\boldsymbol{\alpha}}^*$ and $\hat{\boldsymbol{\beta}}^*$ can be obtained by GNLS as follows:

$$\left(\hat{\boldsymbol{\alpha}}^{*T}, \hat{\boldsymbol{\beta}}^{*T}\right)^T = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\operatorname{argmin}} \sum \left\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{g}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{p}_{\boldsymbol{XZ}})\right\}^T \boldsymbol{\Sigma}^{-1} \left\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{g}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{p}_{\boldsymbol{XZ}})\right\},$$

where $\hat{p}_{\boldsymbol{XZ}}$ is an estimate of $p_{\boldsymbol{XZ}}$ from the IPD.

The diagonal of the covariance matrix $\boldsymbol{\Sigma}$ is typically reported in the literature but the off-diagonals are unknown, thus off-diagonal elements can be imputed by using the IPD. I employ the same imputation method as Debray et al. [50] based on the IPD as follows;

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}_{i,W}) = \hat{V}_i^{\frac{1}{2}} R_{IPD} \hat{V}_i^{\frac{1}{2}},$$

where $\mathrm{Cov}(\hat{\boldsymbol{\theta}}_{i,W})$ is a working covariance matrix of the ith study which is applied to one of the block diagonal elements of $\boldsymbol{\Sigma}$, $\hat{V}_i = \mathrm{diag}(\mathrm{Cov}(\hat{\boldsymbol{\theta}}_i))$ is a diagonal matrix whose diagonal elements are the estimated standard errors (SE) reported from each study and R_{IPD} is a working correlation matrix of coefficients calculated from the IPD. The covariance matrix can be calculated with a sandwich estimator under the model misspecification assumption instead of the imputation based on the IPD [53], but there computational complexity remains a problem and little improvement is gained in simulations studies. Furthermore, even if the covariance matrix is misspecified, the proposed estimator is still consistent and asymptotically normally distributed with a sandwich covariance matrix. This robustness follows the asymptotic theory of the generalized estimating equations. In this situation, let $\hat{\boldsymbol{\alpha}}_W$ and $\hat{\boldsymbol{\beta}}_W$ denote our estimators with the working covariance matrix. The covariance matrix of these estimators can be estimated by

$$\left(\hat{\boldsymbol{D}}^T \boldsymbol{\Sigma}_W^{-1} \hat{\boldsymbol{D}}\right)^{-1} \hat{\boldsymbol{D}}^T \boldsymbol{\Sigma}_W^{-1} \mathrm{Cov}(\hat{\boldsymbol{\theta}}_I) \boldsymbol{\Sigma}_W^{-1} \hat{\boldsymbol{D}} \left(\hat{\boldsymbol{D}}^T \boldsymbol{\Sigma}_W^{-1} \hat{\boldsymbol{D}}\right)^{-1},$$

where $\hat{\boldsymbol{D}} = (\hat{\boldsymbol{D}}_1^T, \ldots, \hat{\boldsymbol{D}}_N^T)^T, \hat{\boldsymbol{D}}_i = \partial \boldsymbol{g}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{p}_{\boldsymbol{XZ}}) / \partial(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)|_{(\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T) = (\hat{\boldsymbol{\alpha}}_W^T, \hat{\boldsymbol{\beta}}_W^T)}$, $\boldsymbol{\Sigma}_W$ is a working covariance matrix, and $\mathrm{Cov}(\hat{\boldsymbol{\theta}}_I) = (\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{g}_i(\hat{\boldsymbol{\alpha}}_W, \hat{\boldsymbol{\beta}}_W, \hat{p}_{\boldsymbol{XZ}})\}\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{g}_i(\hat{\boldsymbol{\alpha}}_W, \hat{\boldsymbol{\beta}}_W, \hat{p}_{\boldsymbol{XZ}})^T\})$ [77, 78]. This idea essentially comes from Liu et al. [78] and can be regarded as analogy of the result proposed by Chen et al. [17].

In addition to the above, if the working covariance matrix is a good approximation of the true covariance structure, the following relationship holds:

$$\mathrm{Avar}(\hat{\boldsymbol{\theta}}^*) \leq \mathrm{Avar}(\hat{\boldsymbol{\theta}}_M) \leq \mathrm{Avar}(\hat{\boldsymbol{\theta}}_S),$$

where Avar denotes an asymptotic covariance matrix, and $\hat{\boldsymbol{\theta}}^*$, $\hat{\boldsymbol{\theta}}_M$ and $\hat{\boldsymbol{\theta}}_S$ are the estimates of $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$ obtained from our proposed method, from multivariate meta-analysis using only the studies with full covariates and from a single study with full covariates, respectively, when the number of studies K goes to infinity.

Here, I assume a fixed effect model which presumes that there is no heterogeneity in the distribution of covariates and in the values of the parameters of interest. This assumption may sometimes be unrealistic. Therefore, I recommend considering whether this assumption is reasonable based on background knowledge or reported information. In addition, I can propose how to modify this to a random effects model to incorporate the heterogeneity by assuming that the parameters underlying studies and the parameters of distribution of covariates follow some distribution. For example, considering the case that all omitted variables are continuous (i.e., section 2.1), I can incorporate random effects by assuming that $\boldsymbol{\alpha}, \boldsymbol{\beta}$, $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}_{\boldsymbol{Z}|\boldsymbol{X}}$ in (4) follow distributions. Random effects in $\boldsymbol{\alpha}, \boldsymbol{\beta}$ accommodate the heterogeneity of parameters and random effects in $\boldsymbol{\Delta}$ and $\boldsymbol{\Omega}_{\boldsymbol{Z}|\boldsymbol{X}}$ accommodates that of distribution of covariates. This is further discussed in the discussion section.

2.4 Special case of logistic regression

2.4.1 The omitted variable bias in the logistic regression model

I introduce the omitted variable bias under one original logistic regression setting, which can afterward be extended to the meta-analysis setting with the assumption that the covariate sets differ among studies.

Let $\boldsymbol{X} = (X_1, \ldots, X_m)$ and $\boldsymbol{Z} = (Z_1, \ldots, Z_l)$ be vectors of covariates and $Y \in \{0, 1\}$ be a binary response variable. Suppose DGP can be formulated by the true model:

$$\text{logit}P(Y = 1|\boldsymbol{X}, \boldsymbol{Z}) = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{Z}\boldsymbol{\beta}, \tag{2.10}$$

where $\boldsymbol{\alpha}, \boldsymbol{\beta}$ are the true parameters of interest and "logit" means the logistic function, $\text{logit}(p) = \log(p/(1-p))$. The misspecified model is assumed to be fitted, which omits relevant covariates $\boldsymbol{Z}$ from the true model Equation 2.10. Specifically,

$$\text{logit}P(Y = 1|\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{\gamma}. \tag{2.11}$$

I investigate the degree to which the regression coefficient $\boldsymbol{\gamma}$, estimated under the misspecified model, differs from the true parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and define the differences as the omitted variable bias.

To derive the omitted variable bias, the unbiasedness condition of the estimating function can be employed [53]. In general, score functions from misspecified models cannot satisfy the unbiasedness condition of estimating functions. Therefore, the first step is to find the solution of the unbiasedness condition of estimating function Equation 2.12, i.e., find $\boldsymbol{\gamma}^* = \boldsymbol{f}(\boldsymbol{\alpha}, \boldsymbol{\beta}, p_{\boldsymbol{XZ}})$, which is the function of the true parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and the joint distribution of covariates, $p_{\boldsymbol{XZ}}$;

$$E\left[\left\{Y - \frac{1}{1 + \exp(-\boldsymbol{X}\boldsymbol{\gamma}^*)}\right\}\boldsymbol{X}^T\right] = 0. \tag{2.12}$$

Here, the expectation is taken by the true joint distribution of Y, $\boldsymbol{X}$ and $\boldsymbol{Z}$ defined from (1) and $p_{\boldsymbol{XZ}}$. Under some regularity conditions, the maximum likelihood estimate of $\boldsymbol{\gamma}$ from the misspecified model Equation 2.11 is a consistent estimate of $\boldsymbol{\gamma}^*$.

Secondly, for assessing biases caused by dropping the important predictors, I assume to have (at least) one IPD with the outcome and the full covariates $\boldsymbol{X}$,

$\boldsymbol{Z}$. This assumption is considered reasonable for researchers who want to develop a new prediction model on their own IPD, incorporating prior summary statistics from regression results. Using the IPD, I can empirically solve Equation 2.12 and derive the omitted variable bias.

Note that in the general case, the function $\boldsymbol{f}$ cannot be written in closed form due to its nonlinearity, but in the following case where every omitted covariate is a continuous variable it can be explicitly written.

2.4.2 Case : Omitted covariates, Z, are continuous variables

In general, the maximum likelihood estimate of $\boldsymbol{\gamma}$ in Equation 2.11 (consistently) estimates $\boldsymbol{\gamma}^*$ as the solution of Equation 2.12. In particular, for the cases of normal continuous variables, the following analytical evaluation can be adapted. Now I suppose $\boldsymbol{Z}|\boldsymbol{X}$ follows the multivariate normal distribution, $N(\boldsymbol{\mu}_{\boldsymbol{Z}|\boldsymbol{X}}, \boldsymbol{\Omega}_{\boldsymbol{Z}|\boldsymbol{X}})$. Based on the normality assumption of $\boldsymbol{Z}|\boldsymbol{X}$, I have

$$\boldsymbol{Z}^T = \boldsymbol{\Delta}\boldsymbol{X}^T + \boldsymbol{\tau}$$

where $\boldsymbol{\Delta} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_l)^T$ is $l \times m$ matrix and $\boldsymbol{\tau} \sim N_{\boldsymbol{\tau}}(\boldsymbol{0}, \boldsymbol{\Omega}_{\boldsymbol{Z}|\boldsymbol{X}})$.

Applying the technique of Chao et al. [79] to our covariate structure and using the probit approximation of logistic distribution, the expectation of Y conditional on $\boldsymbol{X}$ can be expressed as follow:

$$\begin{aligned} E[Y|\boldsymbol{X}] = P(Y=1|\boldsymbol{X}) &= \int \frac{1}{1+\exp\left(-\boldsymbol{X}\boldsymbol{\alpha} - \left(\boldsymbol{\Delta}\boldsymbol{X}^T + \boldsymbol{\tau}\right)^T\boldsymbol{\beta}\right)} N_{\boldsymbol{\tau}}(0, \boldsymbol{\Omega}_{\boldsymbol{Z}|\boldsymbol{X}}) d\boldsymbol{\tau} \\ &\approx \Phi\left[c\left\{\frac{\boldsymbol{X}(\boldsymbol{\alpha} + \boldsymbol{\Delta}^T\boldsymbol{\beta})}{\sqrt{1 + c^2\boldsymbol{\beta}^T\boldsymbol{\Omega}_{\boldsymbol{Z}|\boldsymbol{X}}\boldsymbol{\beta}}}\right\}\right], \end{aligned}$$

where Φ is the cumulative distribution function of standard normal distribution and $c = 16(3)^{1/2}/15\pi$ is the adjustment factor for probit approximation of the logistic distribution proposed by Johnson et al. [80].

In order to satisfy the unbiasedness condition of the estimating function, Equation 2.12, I have

$$E\left[\Phi\left[c\left\{\frac{\boldsymbol{X}(\boldsymbol{\alpha}+\boldsymbol{\Delta}^T\boldsymbol{\beta})}{\sqrt{1+c^2\boldsymbol{\beta}^T\boldsymbol{\Omega}_{\boldsymbol{Z}|\boldsymbol{X}}\boldsymbol{\beta}}}\right\}\right]\boldsymbol{X}^T-\Phi\left\{c(\boldsymbol{X}\boldsymbol{\gamma}^*)\right\}\boldsymbol{X}^T\right]=0$$

Therefore, the function $\boldsymbol{f}$ should be denoted as

$$\boldsymbol{\gamma}^*=\boldsymbol{f}(\boldsymbol{\alpha},\boldsymbol{\beta},p_{\boldsymbol{XZ}})\approx\frac{\boldsymbol{\alpha}+\boldsymbol{\Delta}^T\boldsymbol{\beta}}{\sqrt{1+c^2\boldsymbol{\beta}^T\boldsymbol{\Omega}_{\boldsymbol{Z}|\boldsymbol{X}}\boldsymbol{\beta}}} \tag{2.13}$$

which is the generalization of the results of Chao et al. [79] and Cramer et al. [81].

2.4.3 GLS for synthesis of logistic regression coefficients

As mentioned in the previous section, to synthesize the logistic regression coefficients, the GNLS method can be applied here. The formulation of non-linear model can be written by (which is same with Equation 2.9)

$$\hat{\boldsymbol{\theta}}_i=\boldsymbol{g}_i(\boldsymbol{\alpha},\boldsymbol{\beta},p_{\boldsymbol{XZ}})+\boldsymbol{\varepsilon}_i\qquad(i=1,\ldots,K), \tag{2.14}$$

where

$$\boldsymbol{g}_i(\boldsymbol{\alpha},\boldsymbol{\beta},p_{\boldsymbol{XZ}})=\begin{cases}(\boldsymbol{\alpha}^T,\boldsymbol{\beta}^T)^T & (i=1,\ldots,k)\\ \boldsymbol{f}(\boldsymbol{\alpha},\boldsymbol{\beta},p_{\boldsymbol{XZ}}) & (i=n+1,\ldots,K),\end{cases}$$

$$\begin{pmatrix}\boldsymbol{\varepsilon}_1\\ \vdots\\ \boldsymbol{\varepsilon}_K\end{pmatrix}\sim N(\boldsymbol{0},\boldsymbol{\Sigma}),\quad\boldsymbol{\Sigma}=\begin{bmatrix}\mathrm{Cov}(\hat{\boldsymbol{\theta}}_1) & \ldots & \boldsymbol{0}\\ \vdots & \ddots & \vdots\\ \boldsymbol{0} & \ldots & \mathrm{Cov}(\hat{\boldsymbol{\theta}}_K)\end{bmatrix},$$

and $\hat{\boldsymbol{\theta}}_i$ is the column vector of reported coefficients in the ith study. The function $\boldsymbol{f}()$ comes from the omitted variable bias formula introduced in the previous section, whose formulation is reasonable if an assumption of homogeneity of studies in meta-analysis is acceptable.

By calculating this model by using the technique of GNLS, the synthesized coefficients $\hat{\theta}$ can be obtained. More precisely, overall estimates of the regression coefficients $\hat{\theta}$ (i.e., $\hat{\boldsymbol{\alpha}}^*$ and $\hat{\boldsymbol{\beta}}^*$) can be obtained by finding the point which can minimize the following function:

$$\left(\hat{\boldsymbol{\alpha}}^{*T}, \hat{\boldsymbol{\beta}}^{*T}\right)^T = \underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\operatorname{argmin}} \sum_{i=1}^{N} \left\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{g}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{p}_{\boldsymbol{XZ}})\right\}^T \boldsymbol{\Sigma}^{-1} \left\{\hat{\boldsymbol{\theta}}_i - \boldsymbol{g}_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \hat{p}_{\boldsymbol{XZ}})\right\},$$

where $\hat{p}_{\boldsymbol{XZ}}$ is an estimate of $p_{\boldsymbol{XZ}}$ from the IPD.

2.5 Special case of linear regression

In this case, I can consider the special case where unreported correlation estimates can be recovered only from reported summary statistics. As mentioned in previous section, there are numerous inherent difficulties associated with the individual studies applicable to this area, which include: 1) different sets of covariates exist for the various studies and 2) unreported correlations of covariates exist, which are rarely available in published documentation. In response to the first issue (i.e., different sets of covariates), Resche-Rigon et al. [61] proposed a multiple imputation method for absent covariates with individual patient data (IPD). In addition, the Fibrinogen Studies Collaboration [56] proposed a multivariate meta-analysis approach by using partial models with IPD data obtained from individual studies. These methods have limitations, however, with issues associated with IPD data availability; thus, instead of utilizing IPD, Debray et al. [50] coped with the problem by using mean- or zero-imputation for absent coefficients. However, a bias toward synthesized coefficients is created as the coefficients themselves depend on which covariates are included in the models. To address such a challenge, conventional meta-analysis generally performs multivariate meta-analysis using only studies with common covariate sets. However, this exclusion leads to loss of efficiency because excluded studies may contain indirect information.

The second noted difficulty (i.e., the unreported correlations of covariates) is similar to the problem of within-study correlations in multivariate meta-analysis,

which incorporates the within-study correlations for improving the precision of each estimation compared to those of univariate meta-analyses [41]. Standard multivariate meta-analysis assumes that the within-study correlations are known or IPD are available to borrow estimates of the correlations themselves (for example, applications of the multivariate meta-analysis of rheumatoid arthritis and periodontal data [42]). In the cases where IPD are not available, there are some studies describing the approach that the unknown within-study correlations are: (1) imputed with plausible values per individual professional expertise [82, 83]; (2) imputed with empirical estimations of common Pearson correlations [84]; and (3) imputed by using a Bayesian method assuming prior distributions of correlations [85]. Wei and Higgins [41] provided detailed tables for approximating the within-study correlations based on information regarding likely correlations. However, there are other directly applicable techniques that may be alternatively employed for recovering exact values of unreported covariate correlations from only summary statistics without the use of IPD.

In this section, the synthesis methodology is proposed for regression results under different covariate sets by using a GLS method that includes bias correction terms. In addition, a method for recovering covariate correlations was likewise developed, which is necessary for the implementation of the GLS method. Note that while the synthesis of linear regression models is emphasised, this study can be extended to more complex models, such as generalised linear models.

2.5.1 Setting for synthesis of linear regression results

Let Y denote a dependent variable that is related to independent variables $(X_1, \ldots, X_m, Z_1, \ldots, Z_l)$ and a constant term. I consider a true model (data generating process (DGP)) as follows;

$$Y = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_m X_m + \beta_1 Z_1 + \cdots + \beta_l Z_l + \varepsilon, \tag{2.15}$$

where $\boldsymbol{\varepsilon} \sim N(0, \sigma_\varepsilon^2)$. The misspecified model assumes to omit relevant covariates $\boldsymbol{Z}$ from the fitted model. Specifically,

$$Y = \gamma_0 + \gamma_1 X_1 + \cdots + \gamma_m X_m + \tau, \tag{2.16}$$

where $\tau \sim N(0, \sigma_\tau^2)$. I have N sets of observations on Y and $(X_1, \ldots, X_m, Z_1, \ldots, Z_l)$, which I represent as follows;

$$(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{Z}) = \begin{pmatrix} Y_1 & 1 & X_{11} & \ldots & X_{1m} & Z_{11} & \ldots & Z_{1l} \\ \vdots & \vdots & & \vdots & & & \vdots & \\ Y_N & 1 & X_{N1} & \ldots & X_{Nm} & Z_{N1} & \ldots & Z_{Nl} \end{pmatrix}$$

where X includes 1 in the first column. Our first interest is to derive how much the estimator of regression coefficient $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_m)^T$ differs from that of true parameters $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_m)^T$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_l)^T$, and define the differences as an omitted variable bias. In the context of the meta-analysis of regression coefficients, the concept of omitted variable bias can be regarded analogous to different covariate sets among candidate models under an assumption of homogeneity of studies. For simplicity, we consider the case where the model with full-set covariates (full model) can be defined as a true model. This setting indicates that the prior models are subsets of the full model with underspecified covariate sets. However, this study's method can be characterised into a more general case such that prior models include those with larger sets of covariates than true models.

2.5.2 Omitted variable bias formula in linear regression

This discussion topic is well known and has been assessed by Greene (2003) [77]. The estimator of regression coefficients in (2.16) is provided as the followings;

$$\hat{\boldsymbol{\gamma}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{\alpha} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}\boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon},$$

where we define a sample vector as $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)^T$. By taking the expectation, we can check that unless $\boldsymbol{X}^T\boldsymbol{Z} = 0$ or $\boldsymbol{\beta} = 0$, $\hat{\boldsymbol{\gamma}}$ is biased. In this case, we can

define the omitted variable bias formula:

$$E[\hat{\boldsymbol{\gamma}}] = \boldsymbol{\alpha} + P\boldsymbol{\beta}, \tag{2.17}$$

where $P = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}$.

Another proposed method for deriving this omitted variable bias formula is to utilise the technique of unbiasedness conditions of estimating functions, which can calculate the asymptotic bias for the omission of covariates. This is a general approach and is easy to extend to another model, including generalised linear models. Score functions of misspecified models cannot generally satisfy the unbiasedness conditions of estimating functions. Therefore, the direct approach to calculate the bias is to find $\boldsymbol{\gamma}^* = \boldsymbol{f}(\boldsymbol{\alpha}, \boldsymbol{\beta}, p_{\boldsymbol{XZ}})$ as the solution of the following condition (2.18). In this case, $\boldsymbol{X}$ and $\boldsymbol{Z}$ denote random variables; $\boldsymbol{X} = (1\ X_1 \dots X_m)$ and $\boldsymbol{Z} = (Z_1 \dots Z_l)$. The solution can be written as the function of true parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$ and parameters from joint distribution of covariates, $p_{\boldsymbol{XZ}}$;

$$E\left[\boldsymbol{X}^T\left(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\gamma}^*\right)\right] = \boldsymbol{0}, \tag{2.18}$$

where the expectation is calculated with the true joint distribution of outcome and covariates. The solution, $\boldsymbol{\gamma}^*$, also leads to same omitted variable bias formula (2.17) under the assumption that the joint distribution of covariates follows the multivariate normal distribution, and this case is considered in this study.

2.5.3 GLS for synthesis of linear regression coefficients

Suppose there exist K reported regression models ($j = 1, \dots, K$) with reported coefficients and their covariance matrices. When $j = 1, \dots, k^*$, studies fit the true model with a full set of covariates (full model (1)) $(X_1, \dots, X_m)$, $(Z_1, \dots, Z_l)$ and a constant term, and when $j = k^* + 1, \dots, K$, studies mistakenly fit the model without the covariates (misspecified model (2)) $(Z_1, \dots, Z_l)$. This study only shows the case that $(Z_1, \dots, Z_l)$ is omitted, but the $(X_1, \dots, X_m)$ omitted case can be considered in the same manner. To synthesize the reported coefficient

vectors, the study applied a GLS method to include unequal variances for meta-analysis. This approach was first proposed by Becker and Wu (2007) [33] and our method extends this approach to a more general case where different sets of covariates exist among the candidate models.

Based on these settings, the GLS model for meta-analysis is proposed as follows:

$$\begin{aligned}\hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\boldsymbol{\alpha}}_1 \\ \hat{\boldsymbol{\beta}}_1 \\ \vdots \\ \hat{\boldsymbol{\alpha}}_{k^*} \\ \hat{\boldsymbol{\beta}}_{k^*} \\ \hdashline \hat{\boldsymbol{\gamma}}_{k^*+1} \\ \vdots \\ \hat{\boldsymbol{\gamma}}_K \end{pmatrix} &= \left(\begin{array}{c:c} \multicolumn{2}{c}{\boldsymbol{I}_{(m+l)\times(m+l)}} \\ \multicolumn{2}{c}{\vdots} \\ \multicolumn{2}{c}{\boldsymbol{I}_{(m+l)\times(m+l)}} \\ \hdashline \boldsymbol{I}_{m\times m} & E[(\boldsymbol{X}^T\boldsymbol{X})]^{-1}E[\boldsymbol{X}^T\boldsymbol{Z}] \\ \vdots & \vdots \\ \boldsymbol{I}_{m\times m} & E[(\boldsymbol{X}^T\boldsymbol{X})]^{-1}E[\boldsymbol{X}^T\boldsymbol{Z}] \end{array}\right) \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} + \boldsymbol{e} \\ &= \boldsymbol{W}\boldsymbol{\theta} + \boldsymbol{e} \end{aligned} \tag{2.19}$$

and

$$\boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{Cov}(\hat{\boldsymbol{\theta}}_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \mathrm{Cov}(\hat{\boldsymbol{\theta}}_K) \end{pmatrix}, \tag{2.20}$$

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}}_j) = \begin{cases} \hat{\sigma}_{\varepsilon,j} \begin{pmatrix} \boldsymbol{X}_j^T\boldsymbol{X}_j & \boldsymbol{X}_j^T\boldsymbol{Z}_j \\ \boldsymbol{Z}_j^T\boldsymbol{X}_j & \boldsymbol{Z}_j^T\boldsymbol{Z}_j \end{pmatrix}^{-1} & \text{if } j = 1, \dots, k^* \\ \hat{\sigma}_{\tau,j}(\boldsymbol{X}_j^T\boldsymbol{X}_j)^{-1} & \text{if } j = k^*+1, \dots, K \end{cases} \tag{2.21}$$

where X_j and Z_j are sample covariate matrices at single study j, and $\boldsymbol{I}_{(m+l)\times(m+l)}$ and $\boldsymbol{I}_{m\times m}$ indicate $(m+l)\times(m+l)$ and $m\times m$ identity matrices, respectively. The expectation of $E[(\boldsymbol{X}^T\boldsymbol{X})]^{-1}E[\boldsymbol{X}^T\boldsymbol{Z}]$ can be calculated based on the information obtained from the study fitting the full model with the largest samples. Under the above model, the synthesized estimates of the regression coefficients, $\hat{\boldsymbol{\theta}}^* =$

$(\hat{\boldsymbol{\alpha}}^{*T}, \hat{\boldsymbol{\beta}}^{*T})^T$, and its covariance matrix can be obtained by the GLS method as

$$\hat{\boldsymbol{\theta}}^* = (\boldsymbol{W}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{W})^{-1}\boldsymbol{W}^T\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\theta}} \qquad \text{and} \qquad \text{Cov}(\hat{\boldsymbol{\theta}}^*) = (\boldsymbol{W}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{W})^{-1}.$$

To calculate these estimates, it is necessary to have several summary statistics, such as the covariance matrices: ($\boldsymbol{X}_j^T\boldsymbol{X}_j$, $\boldsymbol{Z}_j^T\boldsymbol{Z}_j$ and $\boldsymbol{X}_j^T\boldsymbol{Z}_j$). However, it is common that although the diagonal elements of covariance matrices (variance of covariates) are frequently reported in literatures, the off-diagonals (covariances) are sometimes unknown. For such cases, it is possible to recover full covariance estimates from only available summary statistics under certain settings, as described in detail in next subsection.

2.5.4 Recover covariance matrix from summary statistics

Let us consider a simple case where there are at most three covariates in a full model, which depicts a model having the highest number of covariates in the meta-analysis can be denoted as:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + e, \quad e \sim N(0, \sigma^2).$$

Under a meta-analysis setting, it is assumed that each study reports the number of sample ($n_j, (j = 1, \ldots, K)$), mean and variance of each covariate ($\bar{X}_{js}$, $\text{Var}(X_{js}), (s = 1, 2, 3)$), an estimated constant term ($\hat{\alpha}_{j0}$), estimated coefficients and its variances ($\hat{\boldsymbol{\alpha}}_j = (\hat{\alpha}_{j1}, \hat{\alpha}_{j2}, \hat{\alpha}_{j3})^T$, $\text{Var}(\hat{\boldsymbol{\alpha}}_j)$) and an estimated standard error of regression ($\hat{\sigma}_j^2$). It is also assumed, however, that correlations among covariates (ρ_{12}, ρ_{13} and ρ_{23}) are not reported. Here we illustrate a methodology to recover these three correlation estimates only from available summary statistics.

To obtain estimates of correlation of covariates, the following formula for the variance of estimated coefficients can be employed. This can be denoted as follows:

$$\text{Var}(\hat{\alpha}_{js}) = \frac{\hat{\sigma}_j^2}{n_j \text{Var}(X_{js})(1 - R_s^2)}, \tag{2.22}$$

where R_s^2 is a coefficient of determination of the regression model of X_{js} on a constant term and other covariates. In this equation, only R_s^2 is unknown and this can be written as the function of the correlation parameters of covariates (the left hand side is assumed to be reported). Since, we have three unknown R_s^2 and three known $\mathrm{Var}(\hat{\boldsymbol{\alpha}}_j)$, it is possible to solve the equations (8) for R_s^2. The proof of this formulation is shown in Appendix D. Intuitively, in this example, the matrix in Equation 2.21 can be denoted as

$$
\begin{aligned}
& \hat{\sigma}_j^2 (\boldsymbol{X}_j^T \boldsymbol{X}_j)^{-1} \\
= \; & \frac{\hat{\sigma}_j^2}{n_j} \begin{pmatrix} \mathrm{Var}(X_{j1}) + \bar{X}_{j1}^2 & \mathrm{Cov}(X_{j1}, X_{j2}) + \bar{X}_{j1}\bar{X}_{j2} & \mathrm{Cov}(X_{j1}, X_{j3}) + \bar{X}_{j1}\bar{X}_{j3} \\ \mathrm{Cov}(X_{j1}, X_{j2}) + \bar{X}_{j1}\bar{X}_{j2} & \mathrm{Var}(X_{j2}) + \bar{X}_{j2}^2 & \mathrm{Cov}(X_{j2}, X_{j3}) + \bar{X}_{j2}\bar{X}_{j3} \\ \mathrm{Cov}(X_{j1}, X_{j3}) + \bar{X}_{j1}\bar{X}_{j3} & \mathrm{Cov}(X_{j2}, X_{j3}) + \bar{X}_{j2}\bar{X}_{j3} & \mathrm{Var}(X_{j3}) + \bar{X}_{j3}^2 \end{pmatrix}^{-1},
\end{aligned}
$$

where $\mathrm{Cov}(X_{js}, X_{js'}) = \rho_{ss'}\sqrt{\mathrm{Var}(X_{js})\mathrm{Var}(X_{js'})}$ is a covariance for the study j. Since three elements in diagonal are assumed known and can be written as the function of ρ_{12}, ρ_{13} and ρ_{23}, we can identify ρ_{12}, ρ_{13} and ρ_{23} by solving the associated equations derived from the diagonal elements.

Chapter 3

Simulations

3.1 Simulation for logistic regression case

3.1.1 Simulation setup for logistic regression case

In this section, I first describe a Monte Carlo simulation which was performed to evaluate the performance of our proposed method for the case of logistic regression. In the simulation, I empirically calculate the omitted variable bias by using equation (3) instead of equation (4). The parameters which varied in the simulation scenario are the true value of a parameter in a DGP model, the number of predictors and the distribution of covariates (continuous/discrete covariates). For simplicity, I examined the case where the number of predictors in the models in this simulation was 1 or 2 (i.e., X_1, X_2 or both). The DGP model was $\text{logit}P(Y = 1|X_1, X_2) = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2$. For checking the sensitivity for the true value of the parameter in the DGP model, α_1 varied from -2 to 1, and true values of other parameters was set at 1 (i.e., $\alpha_0 = \alpha_2 = 1$).

I simulated $N = 9,\ (i = 1, \ldots, 9)$ independent studies with 100 samples in each, and of these studies, 3 studies $(i = 1, 2, 3)$ included a full set of covariates (X_1 and X_2), 3 $(i = 4, 5, 6)$ are supposed to omit X_1 and 3 $(i = 7, 8, 9)$ are supposed to omit X_2. One of the studies with the full set of covariates was used as the IPD. As mentioned above, the off-diagonals of the covariance matrix are

often unknown, thus I adopted the imputation by IPD proposed in the Methods section. In this simulation, I compared the performance of this imputation with the setting using a true covariance structure, which could be estimated from simulation settings.

I classified the scenario into 2 cases according to the distribution of covariates (continuous/discrete distribution). In Case 1, X_1 and X_2 are both continuous and followed the multivariate normal distribution, $X_1, X_2 \sim N\left(\begin{pmatrix}2\\2\end{pmatrix}, \begin{pmatrix}1 & r\\ r & 1\end{pmatrix}\right)$. The correlation, r, between X_1 and X_2 was set at 0 or 0.5. In this case, I checked the performance of the approximation formula (4). Case 2 was the more practical case in which continuous and discrete distributions are mixed (i.e., X_1 was continuous and X_2 was binary). X_2 was binarized from the distribution in Case 1 by a threshold value set at 2. Under these settings, 1000 Monte Carlo simulations are implemented. If the models could not be fitted and converged, their results are excluded from the calculation of bias and mean squared error (MSE).

Performance of the proposed method was evaluated by bias and MSE, comparing it with two ordinary methods. M1 was the multivariate meta-analysis using only 3 studies with a full set of covariates. From a theoretical perspective, the M1 strategy does not include any bias but is inferior in efficiency compared with our proposed method, which can be checked using the results of MSE. M2 was the multivariate meta-analysis after mean imputation of missing coefficients, whose method was proposed in Debray et al. [50]. For example, coefficients and their estimated standard errors of X_2 from 3 studies ($i = 4, 5, 6$) are imputed by the means of the other 6 studies. I tried the zero imputation method, which Debray et al. adopted [50] and called uninformative regression coefficients, but it did not show notable results compared with the results from M2 (mean imputation). Therefore, I decided not to include the results of this method. Note that in this setting, the zero imputation method corresponds to the factored likelihood method proposed by Wu and Becker as explained in the Methods section [51].

3.1.2 Simulation results for logistic regression case

The results of the simulation revealed that compared with the ordinary meta-analysis, our proposed estimator generally produced more precise and less-biased estimates for all simulation settings (Table 3.1 - 3.3). The bias of our estimator ranged from -0.052 to 0.097 (mean: 0.021) for Case 1 and from -0.064 to 0.488 (mean: 0.040) for Case 2. The MSE of our estimator ranged from 0.021 to 0.803 (mean: 0.124) for Case 1 and from 0.012 to 0.486 (mean: 0.091) for Case 2. Although the M2 strategy in Case 1 and $r = 0$ yielded somewhat biased results, the greatest amount of variation seemed to arise from the biased estimates of α_0 in the models from which X_2 was omitted.

The relative efficiency (RE) of the estimates of M1 versus those of our proposed method ranged from 1.023 to 9.913 (mean: 2.323) for Case 1 and from 1.098 to 10.047 (mean: 2.495) for Case 2. The RE of the estimates of M2 ranged from 1.025 to 82.069 (mean: 20.043) for Case 1 and from 0.600 to 93.405 (mean: 123.760) for Case 2.

In terms of the RE of the estimates from the true covariance structure versus the imputation method for unknown elements in the covariance structure, the RE of the covariance structure imputed from the IPD versus the true covariance structure ranged from 0.900 to 1.448 (mean: 1.126) for Case 1 and from 0.895 to 1.193 (mean: 1.065) for Case 2.

Comparing the MSE by correlation value ($r = 0$ versus 0.5), in Case 1 the mean MSE of our proposed method was $r = 0$: 0.074 versus $r = 0.5$: 0.005. In Case 2, the mean MSE of our proposed method was $r = 0$: 0.113 versus $r = 0.5$: 0.174.

TABLE 3.1: Performance of our proposed method on simulation data (Bias) for logistic regression case

	Case 1; Both X_1 and X_2 are continuous								Case 2; X_1 is continuous and X_2 is binary							
	Correlation r=0				Correlation r=0.5				Correlation r=0				Correlation r=0.5			
	Covariance matrix imputed with IPD															
	-2	-1	0	1	-2	-1	0	1	-2	-1	0	1	-2	-1	0	1
Bias	Proposed method															
	α_0 -0.002	0.024	0.030	0.000	0.028	0.029	0.043	0.052	0.025	0.022	0.013	-0.040	0.057	0.037	0.006	-0.006
	α_1 -0.052	-0.022	-0.001	0.086	-0.049	-0.024	-0.011	0.083	-0.050	-0.015	-0.003	0.071	-0.064	-0.017	0.002	0.025
	α_2 0.038	0.022	0.012	0.097	0.030	0.018	0.010	0.058	0.044	0.007	0.018	0.266	0.049	0.001	0.012	0.488
	M1: Full set only															
	α_0 0.011	0.010	-0.018	-0.365	0.009	-0.009	-0.029	-0.058	0.011	0.016	-0.004	-0.016	0.027	0.011	-0.007	-0.002
	α_1 -0.009	0.003	0.011	0.061	0.001	0.002	0.000	0.082	-0.001	0.002	0.000	0.029	-0.006	0.001	-0.004	0.012
	α_2 0.001	-0.009	-0.019	0.061	0.001	0.000	-0.015	0.023	0.008	-0.027	-0.006	0.124	0.027	-0.008	0.001	0.272
	M2: Mean imputation															
	α_0 2.166	1.551	-0.193	-1.275	1.197	0.724	-0.145	-0.590	2.930	2.116	0.048	-1.131	2.612	1.789	0.050	-0.841
	α_1 -1.807	-0.818	-0.171	0.218	-1.921	-1.101	-0.415	-0.135	-1.334	-0.756	-0.082	0.203	-1.385	-0.800	-0.142	0.075
	α_2 0.811	0.218	-0.034	0.226	1.543	0.940	0.109	-0.145	0.034	-0.138	-0.069	-0.460	1.292	0.778	0.010	-1.356
Convergence proportion (%)																
	100	100	99.1	61.5	100	100	98.4	67.3	100	100	98.4	61.2	100	100	98.4	60.3

TABLE 3.2: Performance of our proposed method on simulation data (MSE) for logistic regression case

	Case 1; Both X_1 and X_2 are continuous								Case 2; X_1 is continuous and X_2 is binary							
	Correlation r=0				Correlation r=0.5				Correlation r=0				Correlation r=0.5			
	Covariance matrix imputed with IPD															
	-2	-1	0	1	-2	-1	0	1	-2	-1	0	1	-2	-1	0	1
MSE	Proposed method															
	α_0 0.123	0.098	0.205	0.803	0.085	0.056	0.101	0.251	0.096	0.055	0.065	0.136	0.074	0.044	0.047	0.086
	α_1 0.045	0.021	0.037	0.215	0.060	0.025	0.057	0.181	0.051	0.015	0.012	0.058	0.057	0.017	0.016	0.059
	α_2 0.033	0.022	0.034	0.204	0.044	0.027	0.060	0.188	0.144	0.055	0.051	0.247	0.173	0.070	0.076	0.486
	M1: Full set only															
	α_0 0.248	0.208	0.453	7.960	0.160	0.131	0.274	0.707	0.195	0.116	0.142	0.300	0.169	0.099	0.108	0.197
	α_1 0.068	0.033	0.068	0.620	0.075	0.033	0.094	0.398	0.096	0.027	0.025	0.112	0.099	0.030	0.028	0.112
	α_2 0.038	0.032	0.093	0.591	0.045	0.034	0.118	0.570	0.162	0.077	0.103	2.165	0.190	0.095	0.120	4.883
	M2: Mean imputation															
	α_0 5.290	2.663	0.236	1.929	1.633	0.625	0.113	0.638	8.771	4.532	0.039	1.372	6.912	3.242	0.033	0.775
	α_1 3.694	0.806	0.071	0.242	3.771	1.254	0.214	0.213	2.083	0.601	0.012	0.126	2.075	0.661	0.026	0.087
	α_2 0.870	0.195	0.088	0.209	2.424	0.934	0.091	0.333	0.310	0.098	0.079	2.058	1.919	0.676	0.065	5.396
Convergence proportion (%)																
	100	100	99.1	61.5	100	100	98.4	67.3	100	100	98.4	61.2	100	100	98.4	60.3

TABLE 3.3: Performance of our proposed method on simulation data (Bias and MSE) with true covariance matrix for logistic regression case

		True Covariance matrix															
		-2	-1	0	1	-2	-1	0	1	-2	-1	0	1	-2	-1	0	1
Bias	Proposed method																
	α_0	0.032	0.034	0.049	0.215	0.058	0.038	0.050	0.185	0.059	0.045	0.028	0.032	0.075	0.054	0.014	0.041
	α_1	-0.118	-0.048	-0.005	0.099	-0.110	-0.046	-0.010	0.105	-0.116	-0.037	-0.004	0.072	-0.117	-0.036	0.003	0.036
	α_2	0.070	0.055	0.052	0.104	0.055	0.049	0.058	0.096	0.082	0.036	0.039	0.250	0.072	0.023	0.037	0.462
MSE	Proposed method																
	α_0	0.111	0.089	0.196	0.690	0.077	0.053	0.094	0.204	0.091	0.054	0.065	0.123	0.071	0.043	0.046	0.080
	α_1	0.050	0.021	0.034	0.161	0.061	0.024	0.050	0.125	0.057	0.015	0.012	0.050	0.062	0.017	0.015	0.051
	α_2	0.031	0.022	0.032	0.146	0.039	0.026	0.053	0.136	0.133	0.049	0.049	0.215	0.145	0.060	0.069	0.417
Convergence proportion (%)																	
		100	100	99.1	61.5	100	100	98.4	67.3	100	100	98.4	61.2	100	100	98.4	60.3

3.2 Simulation for linear regression case

3.2.1 Simulation setup for linear regression case

To farther our proposed method, a Monte Carlo simulation was conducted. A case was established where the maximum number of covariates in the simulation's models was three (X_1, X_2 and X_3). The DGP model was $Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \varepsilon$, and the other misspecified models were model (1): $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \tau$, and model (2): $Y = \gamma_0 + \gamma_1 X_1 + \nu$. The true value of α_1 varied from -3 to 3 for sensitivity, while other true values of parameters were fixed at 1 (i.e., $\alpha_0 = \alpha_2 = \alpha_3 = 1$). We made $N = 9$ ($j = 1, \ldots, 9$) artificial studies with 100 samples for each, and of these studies, 3 studies ($i = 1, 2, 3$) were assumed to fit the DGP model, 3 ($i = 4, 5, 6$) were assumed to fit the model (1), and 3 ($i = 7, 8, 9$) were assumed to fit the model (2). Each study was designed to report only summary statistics (means and standard deviations) of covariates, estimates of coefficients, standard errors of coefficients, and standard errors of regression). As mentioned in the previous section, it was also assumed that correlation among covariates was not reported (i.e., the off-diagonal elements of covariance matrices for coefficients was to remain unknown). Therefore, a recovery method was introduced to obtain estimates of covariate correlation, as well as the off-diagonal elements for covariance matrices of estimated coefficients. X_1, X_2 and X_3 followed a multivariate normal distribution, $X_1, X_2, X_3 \sim N\left(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1 & 0.8 & 0.5\\0.8 & 1 & 0.3\\0.5 & 0.3 & 1\end{pmatrix}\right)$. Under these settings, 1000 separate Monte Carlo simulations were implemented. In terms of bias and mean square error (MSE), performance of our proposed method was compared with two conventional evaluation techniques: M1) the multivariate meta-analysis with three studies applying the DGP model and M2) the multivariate meta-analysis method using mean imputation for the absent coefficients in model (1) and model (2). M2 is also applied in Debray et al. [50].

3.2.2 Simulation results for linear regression case

Table 1 shows the results of the Monte Carlo simulation and reveals that our proposed method is superior to previous conventional methods in terms of bias and MSE. The bias of the proposed method ranges from -1.877 to 2.168 (mean: -0.069). This result is significantly superior to the bias of M2 and similar to the bias of M1. Note that M1 does not theoretically include any bias, but it rather loses net efficiency because of the lack of available information due to absent coefficients. This can be verified by performing a check against MSE.

Relative efficiency (RE) between the proposed method and M1 ranges from 1.007 to 2.067 (mean: 1.647), and the RE between the proposed method and M2 ranges from 813.643 to 5803.627 (mean: 3643.837). This result implies that the proposed method can keep the moderate degree of bias comparable to M1, and furthermore improve efficiencies, as compared with the aforementioned conventional methods (M1 and M2). We checked the performance of our method in several different settings, and we got similar results.

TABLE 3.4: Comparison of performance between our proposed and conventional methods on simulation data (Bias and MSE) for linear regression case

Bias		α_1				
		-3	-1	0	1	3
Proposed						
	α_0	-0.012	-0.010	-0.010	-0.006	-0.002
	α_1	-0.015	-0.013	-0.012	-0.012	-0.014
	α_2	0.011	0.011	0.008	0.006	0.010
	α_3	0.015	0.013	0.013	0.013	0.011
M1) Full set only						
	α_0	-0.008	0.002	-0.006	0.004	0.003
	α_1	-0.007	-0.009	-0.002	-0.005	-0.006
	α_2	0.007	0.007	0.004	0.003	0.006
	α_3	0.003	0.001	0.001	0.000	-0.002
M2) Mean imputation						
	α_0	-1.290	-1.330	-1.348	-1.561	-1.746
	α_1	-1.848	-1.844	-1.877	-1.679	-1.578
	α_2	0.823	0.755	0.409	0.997	0.734
	α_3	1.544	1.635	2.168	1.314	1.593

MSE		α_1				
		-3	-1	0	1	3
Proposed						
	α_0	0.106	0.106	0.104	0.108	0.107
	α_1	0.172	0.177	0.175	0.175	0.174
	α_2	0.133	0.137	0.134	0.135	0.134
	α_3	0.118	0.117	0.116	0.118	0.116
M1) Full set only						
	α_0	0.213	0.216	0.213	0.217	0.222
	α_1	0.300	0.303	0.294	0.297	0.300
	α_2	0.247	0.249	0.243	0.248	0.245
	α_3	0.119	0.118	0.117	0.119	0.117
M2) Mean imputation						
	α_0	528.634	490.720	546.136	438.223	528.150
	α_1	163.996	156.515	194.365	142.313	174.343
	α_2	484.841	435.885	447.087	431.122	525.674
	α_3	636.938	584.995	612.789	641.022	672.318

Chapter 4

Real Data Analysis

4.1 Application in risk prediction models for occurrences of stroke

I applied the proposed methods for a series of epidemiologic studies that developed risk prediction models for occurrences of stroke. Stroke is one of the leading causes of death or physical/cognitive impairment both in developed and developing countries and, therefore, numerous prediction model and clinical characteristics have been modeled and identified as potential predictors [86, 87]. However, overall relationship is still unclear with conflicting results from several literatures [87].

4.2 Application setup

I obtained 10 IPDs conducted by the Japan Public Health Center-based Prospective Study (JPHC study). JPHC study is 30 year on-going prospective cohort and covers 11 public health center areas (Area 1 - 11) across Japan. The JPHC study was initiated by four cohorts and a check-up cohort (Cohort 1) in 1990 and Cohort 2 was included with seven cohorts in 1993. The detail of the location

of study cohort is shown in Figure 4.1, which is adopted from [88] with small revision.

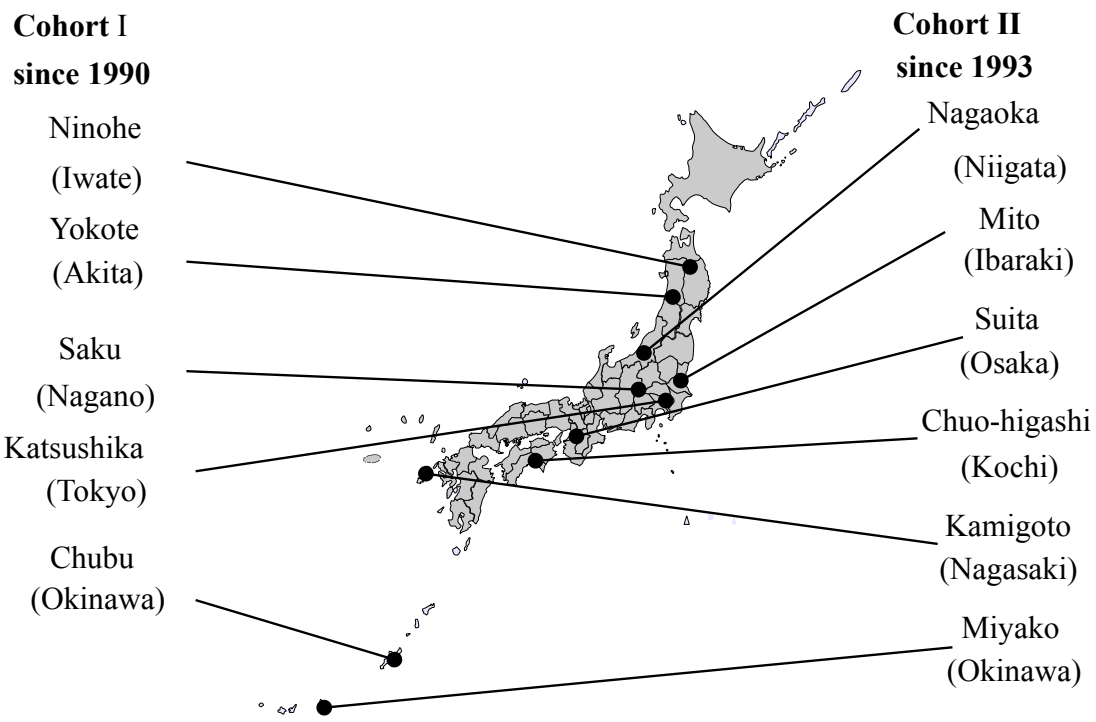

FIGURE 4.1: Map of JPHC cohorts

Total number of participants was 140, 420 and population of the study was the residents who are 40 to 69 years old at the time of the baseline survey. The brief time schedule of the JPHC study is shown in Figure 4.2 with small revision from [88].

Details of study design are well documented in previous study [89]. The outcome was confirmed according to the criteria provided by the National Survey of Stroke, which required a constellation of neurological deficits of sudden or rapid onset lasting at least 24 hours or until death [90, 91].

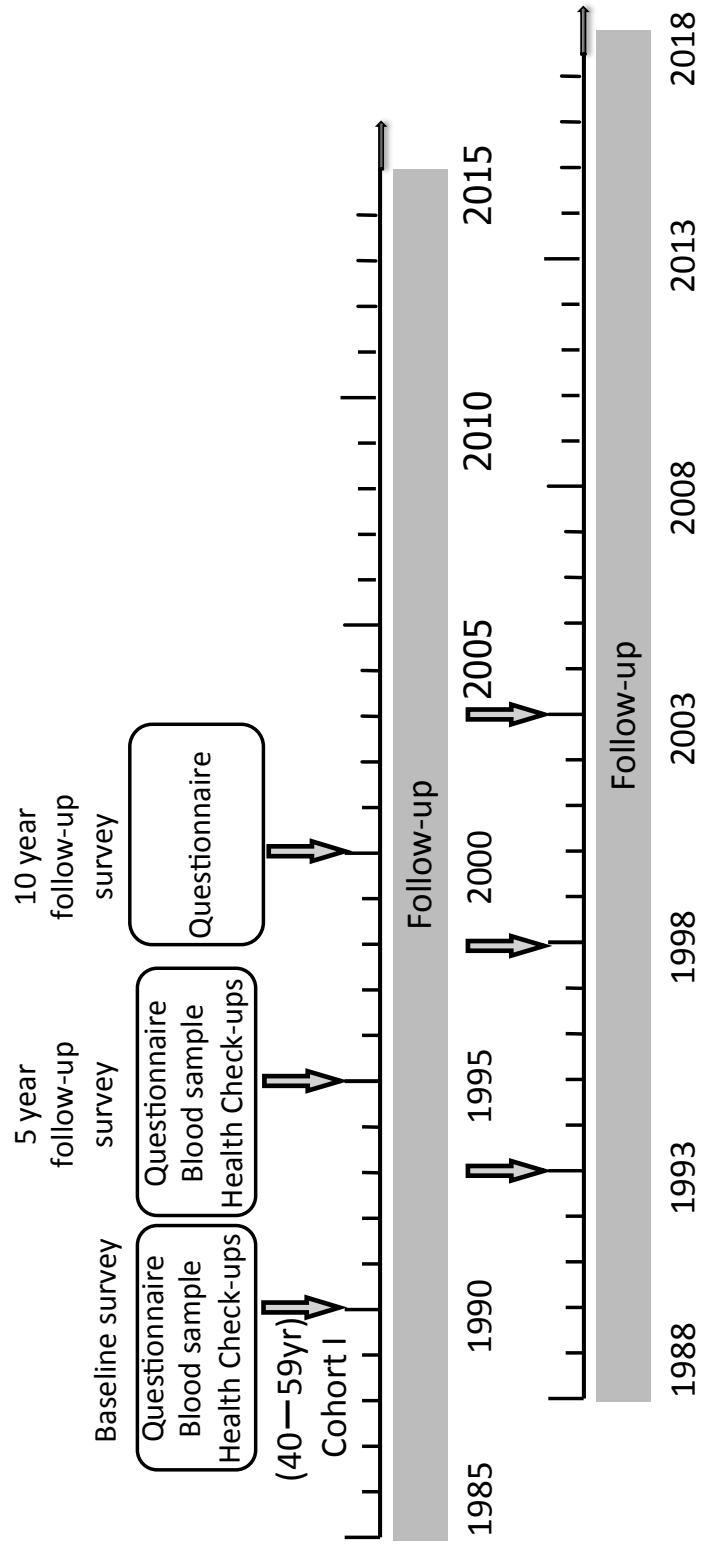

FIGURE 4.2: Schedule of JPHC study

I fitted a logistic regression model to each available IPD and explored the important covariates related to patient characteristics and metabolic syndrome such as age (year), time since last meal (minutes), body mass index; BMI (kg/m^2), total cholesterol level (mg/dl), blood pressure (mmHg), cigarettes (per day), diabetes (yes/no), blood glucose (mg/dl), high-density lipoprotein; HDL (mg/dl), serum triglycerides (mg/dl) (Table 2) [92, 93, 94]. The sets of available covariates are different by region. For example, IPD from Area 1 cohort did not include data of blood glucose, HDL, serum triglycerides since subjects in that cohort did not take any blood test. To overcome this discrepancy among cohorts, which can be typically found in such large scale cohort study collected for several outcomes, was one of our motivation of this study.

Coefficients from each model are stored as the aggregated statistics, which could be regarded as prior studies for meta-analysis. In terms of handling sporadically missing data (average missing rate was 2.8% and standard deviation was 2.5%), complete case analysis was executed. One cohort (Area 9) remained as IPD and one cohort (Area 11) was used as a test data for prediction. Next, I compared our methodology with conventional multivariate meta-analysis using only studies with a full set of covariates and with results from IPD data only.

Lastly, new prediction models are constructed by plugging the synthesized coefficients into the model and check its performance in the test data.

The discriminant performance of the prediction models was measured by the area under the receiver operator characteristic curve (AUC) and the Brier score (BS) (multiplied by 100), which are one of indicators of accuracy of the prediction model. Higher AUC means higher prediction accuracy and BS is vice versa [95]. In addition, the model's calibration was examined by a Hosmer-Lemeshow's chi-squared-statistic [96].

4.3 Application results

The result demonstrated that our approach provided considerably narrower confidence intervals and slightly better prediction performance compared with conventional multivariate meta-analysis (Table 4.1 and Table 4.2). Our estimator reduced SE by 38%-53% and 56%-71% compared with SE from conventional meta-analysis and from IPD, respectively.

According to the result of prediction performance, the prediction model constructed from the synthesized coefficients showed the slight improvements, particularly in BS. AUC and BS are increased by 1.1% and -1.0% on average compared with conventional meta-analysis and decreased by -0.4% and 1.0% on average compared with IPD. The improvements in prediction performance are relatively small because the cohort of test data and other cohorts aggregated into summary statistics as previously published studies are remarkably similar across Japan.

TABLE 4.1: Estimated regression coefficients (and standard error) from JPHC data

	Area 1	Area 2	Area 3	Area 4	Area 5	Area 6
Sample	2121	1678	3396	859	3135	538
Incedence of stroke	109	82	132	23	142	35
Intercept	-12.280 (1.394)	-9.828 (1.668)	-9.917 (1.225)	-8.703 (2.846)	-11.940 (1.367)	-9.475 (3.070)
Age	0.085 (0.022)	0.095 (0.024)	0.066 (0.018)	0.022 (0.042)	0.113 (0.016)	0.126 (0.042)
Postprandial time	-0.016 (0.026)	-0.023 (0.041)	-0.019 (0.018)			
BMI	0.001 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.002)	0.000 (0.001)	-0.002 (0.002)
Total cholesterol level	0.003 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.006)	-0.004 (0.003)	-0.011 (0.006)
Blood pressure	0.027 (0.005)	0.016 (0.007)	0.028 (0.005)	0.022 (0.012)	0.020 (0.006)	0.013 (0.010)
Smoke (per day)	0.020 (0.009)	0.028 (0.010)	0.010 (0.007)	0.012 (0.019)	0.000 (0.007)	-0.019 (0.021)
Diabetes	0.397 (0.504)	1.202 (0.525)	0.738 (0.362)	0.240 (1.314)	0.302 (0.314)	-0.174 (1.065)
Glucose			-0.004 (0.004)	0.012 (0.006)	0.004 (0.003)	
HDL					-0.005 (0.007)	0.012 (0.014)
Triglycerides			0.001 (0.001)		0.000 (0.001)	0.001 (0.003)
AUC	67.01	68.74	67.97	65.52	69.16	68.19
Brier score	7.71	7.72	7.65	8.07	7.78	7.68
Hosmer-Lemeshow	10.91	15.78*	14.01	101.45*	54.96*	17.70*

Proposed: our proposed method; Conventional: conventional meta-analysis using only studies with a full set of covariates. Area 9 is IPD.
*: p-value of Hosmer-Lemeshow test is less than 0.05.

Table 4.2: Cont. Estimated regression coefficients (and standard error) from JPHC data

	Area 7	Area 8	Area 9	Area 10	Proposed	Conventional
Sample	1601	1731	1586	2725		
Incedence of stroke	85	90	90	52		
Intercept	-9.223 (1.710)	-8.413 (1.499)	-10.300 (1.729)	-10.500 (1.878)	-10.170 (0.633)	-9.408 (0.989)
Age	0.088 (0.020)	0.072 (0.018)	0.096 (0.021)	0.069 (0.020)	0.067 (0.007)	0.060 (0.011)
Postprandial time	-0.009 (0.024)	-0.007 (0.025)	-0.006 (0.019)	0.018 (0.034)	0.013 (0.011)	0.017 (0.013)
BMI	-0.001 (0.001)	-0.002 (0.001)	-0.001 (0.001)	0.000 (0.001)	0.000 (0.000)	0.000 (0.001)
Total cholesterol level	0.001 (0.004)	0.001 (0.003)	0.001 (0.004)	-0.003 (0.005)	-0.001 (0.001)	0.001 (0.002)
Blood pressure	0.011 (0.006)	0.015 (0.006)	0.015 (0.007)	0.017 (0.007)	0.017 (0.002)	0.011 (0.004)
Smoke (per day)	0.025 (0.010)	0.007 (0.010)	0.020 (0.009)	0.011 (0.010)	0.013 (0.004)	0.020 (0.006)
Diabetes	0.168 (0.455)	0.052 (0.485)	0.268 (0.490)	0.694 (0.465)	0.158 (0.180)	0.084 (0.262)
Glucose	0.009 (0.003)		0.004 (0.004)	-0.001 (0.008)	0.010 (0.001)	0.014 (0.002)
HDL	-0.022 (0.010)	-0.001 (0.009)	-0.013 (0.010)	0.005 (0.012)	-0.004 (0.005)	-0.008 (0.006)
Triglycerides	-0.003 (0.002)	-0.002 (0.002)	-0.001 (0.001)	0.002 (0.002)	0.000 (0.001)	0.000 (0.001)
AUC	68.32	69.47	68.29	67.28	68.01	67.24
Brier score	7.77	7.63	7.64	8.03	7.72	7.80
Hosmer-Lemeshow	58.60*	28.70*	25.38*	186.76*	21.13*	21.17*

Proposed: our proposed method; Conventional: conventional meta-analysis using only studies with a full set of covariates. Area 9 is IPD.
*: p-value of Hosmer-Lemeshow test is less than 0.05.

Chapter 5

Discussions

Along with increasing attention to prediction models, there has been higher demand for approaches to the meta-analysis of regression coefficients. However these methodologies are not well developed due to the many difficulties caused by the different settings used by various studies, and further research is still needed, particularly compared with conventional meta-analysis methods such as synthesizing mean differences, correlation and so on [33]. This study demonstrated a method to conduct the meta-analysis of regression coefficients with different covariate sets under the assumption of homogeneity of studies (i.e., it is applicable in cases where studies in the meta-analysis have similar distributions of covariates and outcomes). Although this study temporarily assumed the models with a full set of covariates as a true model, our approach can be generalized to any formulation of previous models even if they are over-/under-specified compared to a constructing model. We notice, however, that we need careful arguments about what is an appropriate covariate set. Further, the assumption that (at least) one IPD is available can be considered reasonable in the frequent case in which a single researcher wants to construct a new prediction model on his or her own IPD, incorporating prior regression results (but with such prior results reported just in the form of summary statistics). The minimal use of IPD (use of one IPD and other summary statistics) distinguishes our approach from that of the Fibrinogen Studies Collaboration [56]. They assume that both full and partial models are applied in each cohort by using its cohort IPD, and thus

the estimation of the correlation of coefficients between full and partial models is applicable. Regarding these discussions, this study can provide the following guidelines for practitioners about how to analyze prior models with their own IPD by recognizing the issue of omitted variable bias as the differences of sets of predictors between their constructing models and prior models: 1) the first step is to construct a new and temporal model on their own data set, and 2) the second step is to apply our method to synthesize the previous regression coefficients with their temporal model and then update the model and obtain more accurate estimators.

Our method proved robust against the misspecification of the covariance structure. Because of this property we can arbitrarily set the covariance matrix of coefficients and thus it is possible to avoid the argument, often discussed with methods such as that of Becker and Wu [33], on whether the full covariance matrix of coefficients should be reported or not. This robustness property can be considered as an analogical result provided by Liu et al. [78]. They provide a framework of meta-analysis under heterogeneity by using a confidence density function and reparametrization of the problem setting. Their approach utilizes the reparameterization connecting each study-specific parameter to the common parameter using the transformation function $\boldsymbol{M}_i$, which is used as the omitted variable bias formula in our setting. However, they assume that the omitted covariates are fixed values and thus they can estimate $\boldsymbol{M}_i$ without a consideration of the distribution of covariates. In contrast, our approach provides more general guidelines for treating missing covariates in the meta-analysis.

The simulation performed in this study illustrated that our method is unbiased and has greater efficiency than a conventional meta-analysis approach as well as the technique proposed by Debray et al. [50]. Although our estimator was most efficient if the covariance structure was truly specified, it maintained its efficiency even if we misspecified the covariance structure, with a loss of efficiency by misspecification of only around 10%.

Finally, we demonstrated the practical use of our approach with medical data on stroke prediction. Although the improvement of accuracy of the prediction model was relatively small, the confidence intervals of synthesized coefficients

were dramatically decreased because information from other studies helped improve efficiency. This result can be considered as one of extension of prior results related with methodological studies regarding multivariate meta-analysis. In the context of multivariate meta-analysis, it is well known that we can gain precision by borrowing strength from other partially reported results [82, 16, 19]. This implies that our methodology can be applied not only to prediction models but also to observational studies such as a case-control/cross-sectional study whose main purpose is to identify causal effects.

5.1 Limitations

As a limitation of this study, our method was examined in only one practical dataset. Although this data includes over 100,000 samples, the population was Japanese only, and can thus be regarded as one group with small heterogeneity. This situation may not be representative of an ordinary meta-analysis because the majority of recent meta-analyses include several groups with large heterogeneity due to studies undertaken globally. We think, however, that we took this heterogeneity into account by incorporating random effects, as mentioned in the Methods section. We welcome the re-evaluation of our method in other practical cases.

Another potential limitation is that we implicitly assumed that the distributions of covariates are (approximately) the same between studies. This assumption can also be relaxed by incorporating random effects into parameters related to the distribution, as discussed in the Methods section. However, a random effect model obscures the objective of a meta-analysis because under this model, a global average effect and the effect prevailing in particular circumstances are not identical [40]. We need further research about how to incorporate random effects and its interpretation.

Furthermore, an approach to recover correlation estimates in the case of three covariates was presented in the section 2.5.4. The rationale behind the

restricted number of covariates is that only three estimated variances of coefficients are available to recover three correlations of covariates. If the number of covariates are to be four, then we need to recover ${}_4C_2 = 6$ correlation estimates from four reported variances of coefficients, but this is an indeterminate scenario. However, it should be noted that it is possible to recover over three estimates of correlations by combining subset results under the assumption of homogeneity for the distribution of covariates. For example, the correlations between X_1, X_2, X_3 and X_4 can be calculated under the assumption of homogeneity of studies, if there are two subset models including X_1, X_2 and X_1, X_3. In such a case, we can recover the correlations by using the combinations of reported summary statistics from these studies.

5.2 Future studies

As the future studies, I am considering several extensions of this study and comparisons with other similar techniques such as the synthesis methods using IPD (i.e., IPD meta-analysis).

Firstly, as mentioned in the previous sections, this study assumes that there are no heterogeneity among studies in meta-analysis, indicating that the model is fitted as fixed effect meta-analysis model. For example to extend this fixed effect model into random effect model, I can propose two extensions in the logistic regression case. First extension is related to the parameters of interest in Equation 2.14. Equation 2.14 can be easily extended to random effect model as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_i &= \boldsymbol{g}_i(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i, p_{\boldsymbol{XZ}}) + \boldsymbol{\varepsilon}_i \qquad (i = 1, \ldots, N) \\ \boldsymbol{\theta}_i &= \begin{pmatrix} \boldsymbol{\alpha}_i \\ \boldsymbol{\beta}_i \end{pmatrix} \sim N\left(\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix}, \boldsymbol{V}_1 \right), \end{aligned}$$

where $\boldsymbol{V}_1$ can be considered as the between study covariance matrix in the context of multivariate meta-analysis and $\boldsymbol{\theta}$ is of interest and the average global effect. Second extension can be considered in the expectation operator in Equation 2.12. This expectation is calculated with respect to the true distribution, implicating

that there is one true distribution among studies even if studies were conducted in different population. Therefore, one idea for the extension for the random effect model makes the true distribution vary based on each study's population. For example, in the case of logistic regression and that the omitted covariates are continuous, Equation 2.13 can be extended to include the random effect as

$$\begin{aligned} \boldsymbol{\gamma}^* &= \boldsymbol{f}(\boldsymbol{\alpha}, \boldsymbol{\beta}, p_{i,\boldsymbol{XZ}}) \approx \frac{\boldsymbol{\alpha} + \boldsymbol{\Delta}_i^T \boldsymbol{\beta}}{\sqrt{1 + c^2 \boldsymbol{\beta}^T \boldsymbol{\Omega}_{\boldsymbol{Z_i}|\boldsymbol{X_i}} \boldsymbol{\beta}}} \\ \boldsymbol{\Delta}_i &\sim N(\boldsymbol{\Delta}, \boldsymbol{V}_2) \\ \boldsymbol{\Omega}_{\boldsymbol{Z_i}|\boldsymbol{X_i}} &\sim W(\Psi, v), \end{aligned}$$

where W is a inverse Wishart distribution with the matrix of scale parameters Ψ and the degree of freedom v.

Second future work is to compare the method proposed in this study with the IPD meta-analysis method such as [56]. [56] tackled to the same problem I considered in this study and they studied in the case where IPD from each study are available. By comparison of my method to their method (which should be most efficient because of the availability of IPD), the difference in efficiency between my method and gold standard (IPD meta-analysis) would be clarified.

Finally, this study focused on the improvement of efficiency compared with the ordinary methods, but in terms of predictive performance, it would be useful to develop the methods for synthesizing prediction models to improve the prediction compared with a single prediction model. For understand and further development of this synthesis method, techniques and studies in the field of the machine learning such as a transfer learning [97] (also called a multi-task learning [98] or a learning to learn [99]) and boosting would be helpful.

Chapter 6

Conclusions

This study proposed a correction method for the omitted variable bias due to different sets of covariates between literature models in meta-analysis and our approach and nonlinear models for meta-analysis to borrow strength from misspecified models by using the omitted variable bias formula. By both simulation and theory, it is proved that our method can attain the efficiency compared with the conventional approach. Further, this study also provides a recover method of correlations statistics without IPD for applying the GLS method to synthesize the regression results. This study should be useful for practitioners who want to develop their prediction model on their own dataset with incorporating the prior regression results.

Appendix A

Simulation codes

In this appendix, I provide some example code written in R language for a certain simulation illustrated in simulation section.

zw

```
############################
#This codes were written by Daisuke Yoneoke
#First revise: Sep 13 2013
#Last revise: Jan 17 2015
############################

#########################
#General information
#Study 1~3 have
#       3 parameter (1 intercept and 2 slopes)
#       100 sample

#Study 4~6 have
#       2 parameter (1 intercept and 1 slope)
#       100 sample

#Study 7~9 have
#       2 parameter (1 intercept and 1 slope)
#       100 sample

#########################
library(MASS)
library(Matrix)
library(nlme)
library(glm2)
library(mvmeta)
library(nleqslv)
```

```
#Set basic statics ##################################
#latent data generating process
set.seed(123)

#Make true coefs
beta1<-c(-3,-1,0,1,3)
beta<-rep(1,2)

c<-16*sqrt(3)/15/pi

#Correlation
cor12<-0.5

#Variance
var1<-1
var2<-1

#mean
mu<-c(0,0)

#Make covariance matrix
Xdiag<-matrix(c(
        var1, 0,
        0,var2
        ),nrow=2)

R<-matrix(c(
        1, cor12,
        cor12,1
        ),nrow=2)

X<-Xdiag%*%R%*%Xdiag

sampling<-function(N,mu,X,j){
        data<-mvrnorm(N,mu=mu,Sigma=X)
        data[,2]<-ifelse(data[,2]>0,1,0)

        data2<-as.data.frame(cbind(
        Y=rbinom(N,1,1/(1+exp(-beta[1]-beta1[j]*data[,1]-beta[2]*
   data[,2]))),
        data
        ))
        colnames(data2)<-c("Y","X1","X2")
        return(data2)
}
```

```
#Make sample studies
MakeSample<-function(N1,N2,N3,N4,N5,N6,N7,N8,N9,mu,X,j){
        #Make sample population
        sample1<-sampling(N1,mu,X,j)
        sample2<-sampling(N2,mu,X,j)
        sample3<-sampling(N3,mu,X,j)
        sample4<-sampling(N4,mu,X,j)
        sample5<-sampling(N5,mu,X,j)
        sample6<-sampling(N6,mu,X,j)
        sample7<-sampling(N7,mu,X,j)
        sample8<-sampling(N8,mu,X,j)
        sample9<-sampling(N9,mu,X,j)

        #True models
        result1<-glm2(sample1$Y~X1+X2,data=sample1,family=binomial
    )
        result2<-glm2(sample2$Y~X1+X2,data=sample2,family=binomial
    )
        result3<-glm2(sample3$Y~X1+X2,data=sample3,family=binomial
    )

        result4<-glm2(sample4$Y~X1,data=sample4,family=binomial)
        result5<-glm2(sample5$Y~X1,data=sample5,family=binomial)
        result6<-glm2(sample6$Y~X1,data=sample6,family=binomial)

        result7<-glm2(sample7$Y~X2,data=sample7,family=binomial)
        result8<-glm2(sample8$Y~X2,data=sample8,family=binomial)
        result9<-glm2(sample9$Y~X2,data=sample9,family=binomial)


        if(result1$converged==TRUE & result2$converged==TRUE &
    result3$converged==TRUE & result4$converged==TRUE & result5$
    converged==TRUE & result6$converged==TRUE & result7$converged==
    TRUE & result8$converged==TRUE & result9$converged==TRUE ){
                cov_b1<-vcov(result1)
                delta1 <- row(cov_b1) - col(cov_b1)
                cov_b2<-vcov(result2)
                delta2 <- row(cov_b2) - col(cov_b2)
                cov_b3<-vcov(result3)
                delta3 <- row(cov_b3) - col(cov_b3)
                cov_b4<-vcov(result4)
                delta4 <- row(cov_b4) - col(cov_b4)
                cov_b5<-vcov(result5)
                delta5 <- row(cov_b5) - col(cov_b5)
                cov_b6<-vcov(result6)
```

```
                delta6 <- row(cov_b6) - col(cov_b6)
                cov_b7<-vcov(result7)
                delta7 <- row(cov_b7) - col(cov_b7)
                cov_b8<-vcov(result8)
                delta8 <- row(cov_b8) - col(cov_b8)
                cov_b9<-vcov(result9)
                delta9 <- row(cov_b9) - col(cov_b9)

                lsigma<-as.matrix(bdiag(cov_b1,cov_b2,cov_b3,cov_
   b4,cov_b5,cov_b6,cov_b7,cov_b8,cov_b9))


                return(list(c(result1$coefficients,result2$
   coefficients,result3$coefficients,result4$coefficients,result5$
   coefficients,result6$coefficients,result7$coefficients,result8$
   coefficients,result9$coefficients),lsigma,sample1))
        }else{
                return(NA)
        }

}




W<-as.data.frame(matrix(c(
        1,0,0,0,0,0,0,
        0,1,0,0,0,0,0,
        0,0,1,0,0,0,0,

        1,0,0,0,0,0,0,
        0,1,0,0,0,0,0,
        0,0,1,0,0,0,0,

        1,0,0,0,0,0,0,
        0,1,0,0,0,0,0,
        0,0,1,0,0,0,0,


        0,0,0,1,0,0,0,
        0,0,0,0,1,0,0,

        0,0,0,1,0,0,0,
        0,0,0,0,1,0,0,

        0,0,0,1,0,0,0,
        0,0,0,0,1,0,0,
```

```
        0,0,0,0,0,1,0,
        0,0,0,0,0,0,1,

        0,0,0,0,0,1,0,
        0,0,0,0,0,0,1,

        0,0,0,0,0,1,0,
        0,0,0,0,0,0,1

),ncol=7,byrow=T))
```

```
unbiasx<-function(par.a,par.b,sample){
        a0<-par.a[1]
        a1<-par.a[2]
        a2<-par.a[3]

        b0<-par.b[1]
        b1<-par.b[2]

        Ux<-mean(1/(1+exp(-a0-a1*sample[,2]-a2*sample[,3]))-1/(1+
   exp(-b0-b1*sample[,2])))
        Lx<-mean(sample[,2]*(1/(1+exp(-a0-a1*sample[,2]-a2*sample
   [,3]))-1/(1+exp(-b0-b1*sample[,2]))))
        return(c(Ux,Lx))
}

unbiasz<-function(par.a,par.b,sample){
        a0<-par.a[1]
        a1<-par.a[2]
        a2<-par.a[3]

        b0<-par.b[1]
        b2<-par.b[2]

        Uz<-mean(1/(1+exp(-a0-a1*sample[,2]-a2*sample[,3]))-1/(1+
   exp(-b0-b2*sample[,3])))
        Lz<-mean(sample[,3]*(1/(1+exp(-a0-a1*sample[,2]-a2*sample
   [,3]))-1/(1+exp(-b0-b2*sample[,3]))))
        return(c(Uz,Lz))
}
```

```
res.unbiasx<-function(par.a){
        res<-nleqslv(x=c(1,1),function(x) unbiasx(par.a=par.a,par.
    b=x,sample=sample))
        return(res)
}

res.unbiasz<-function(par.a){
        res<-nleqslv(x=c(1,1),function(x) unbiasz(par.a=par.a,par.
    b=x,sample=sample))
        return(res)
}

resfun <- function(par.a){
        yhat<-par.a[1]*W2$V1+par.a[2]*W2$V2+par.a[3]*W2$V3+
                        W2$V4*res.unbiasx(par.a)$x[1]+W2$V5*res.
    unbiasx(par.a)$x[2]+
                        W2$V6*res.unbiasz(par.a)$x[1]+W2$V7*res.
    unbiasz(par.a)$x[2]

        return(t(as.vector(W2$coef-yhat))%*%solve(Sigma)%*%as.
    vector((W2$coef-yhat)))
}


#Set the matrix to put the results
bias0<-matrix(0,1000,ncol=5)
bias1<-matrix(0,1000,ncol=5)
bias2<-matrix(0,1000,ncol=5)

bias0sub<-matrix(0,1000,ncol=5)
bias1sub<-matrix(0,1000,ncol=5)
bias2sub<-matrix(0,1000,ncol=5)

beta_t00<-matrix(0,1000,ncol=5)
beta_t01<-matrix(0,1000,ncol=5)
beta_t02<-matrix(0,1000,ncol=5)

beta_t10<-matrix(0,1000,ncol=5)
beta_t11<-matrix(0,1000,ncol=5)
beta_t12<-matrix(0,1000,ncol=5)



#Main loop
for (j in 1:5){
        i<-1
```

```
     while (i <= 1000){
             coefs<-MakeSample
(100,100,100,100,100,100,100,100,100,mu,X,j)
             sample<-coefs[[3]]
             cormat<-cov2cor(vcov(glm2(sample$Y~.,data=as.data.
frame(sample),family=binomial)))

             if(is.na(coefs[1])==FALSE){

                     data<-as.data.frame(rbind(
                             coefs[[1]][1:3],
                             coefs[[1]][4:6],
                             coefs[[1]][7:9],

                             c(coefs[[1]][10:11],mean(coefs
[[1]][c(3,6,9,17,19,21)])),
                             c(coefs[[1]][12:13],mean(coefs
[[1]][c(3,6,9,17,19,21)])),
                             c(coefs[[1]][14:15],mean(coefs
[[1]][c(3,6,9,17,19,21)])),

                             c(coefs[[1]][16],mean(coefs[[1]][c
(2,5,8,11,13,15)]),coefs[[1]][17]),
                             c(coefs[[1]][18],mean(coefs[[1]][c
(2,5,8,11,13,15)]),coefs[[1]][19]),
                             c(coefs[[1]][20],mean(coefs[[1]][c
(2,5,8,11,13,15)]),coefs[[1]][21])
                     ))


                     data1<-data[1:3,]

                     #Impute the off-diagonal of cov matrix of
 coefficient in full sets
                     cov1<-matrix(0,ncol=3,nrow=3)
                     diag(cov1)<-sqrt(diag(coefs[[2]][c(1:3),c
(1:3)]))
                     cov1<-cov1%*%cormat%*%cov1


                     cov2<-matrix(0,ncol=3,nrow=3)
                     diag(cov2)<-sqrt(diag(coefs[[2]][c(4:6),c
(4:6)]))
                     cov2<-cov2%*%cormat%*%cov2

                     cov3<-matrix(0,ncol=3,nrow=3)
                     diag(cov3)<-sqrt(diag(coefs[[2]][c(7:9),c
(7:9)]))
```

```
                    cov3<-cov3%*%cormat%*%cov3

                    #Impute the off-diagonal of cov matrix of
 coefficient in omitted sets
                    cov4<-matrix(0,ncol=3,nrow=3)
                    diag(cov4)<-sqrt(c(diag(coefs[[2]][c
(10,11),c(10,11)]),0))
                    cov4[3,3]<-sqrt(mean(cov1[3,3],cov2[3,3],
cov3[3,3]))
                    cov4<-cov4%*%cormat%*%cov4

                    cov5<-matrix(0,ncol=3,nrow=3)
                    diag(cov5)<-sqrt(c(diag(coefs[[2]][c
(12,13),c(12,13)]),0))
                    cov5[3,3]<-sqrt(mean(cov1[3,3],cov2[3,3],
cov3[3,3]))
                    cov5<-cov5%*%cormat%*%cov5

                    cov6<-matrix(0,ncol=3,nrow=3)
                    diag(cov6)<-sqrt(c(diag(coefs[[2]][c
(14,15),c(14,15)]),0))
                    cov6[3,3]<-sqrt(mean(cov1[3,3],cov2[3,3],
cov3[3,3]))
                    cov6<-cov6%*%cormat%*%cov6

                    cov7<-matrix(0,ncol=3,nrow=3)
                    diag(cov7)<-sqrt(c(diag(coefs[[2]])[c(16)
],0,diag(coefs[[2]])[c(17)]))
                    cov7[2,2]<-sqrt(mean(cov1[2,2],cov2[2,2],
cov3[2,2]))
                    cov7<-cov7%*%cormat%*%cov7

                    cov8<-matrix(0,ncol=3,nrow=3)
                    diag(cov8)<-sqrt(c(diag(coefs[[2]])[c(18)
],0,diag(coefs[[2]])[c(19)]))
                    cov8[2,2]<-sqrt(mean(cov1[2,2],cov2[2,2],
cov3[2,2]))
                    cov8<-cov8%*%cormat%*%cov8

                    cov9<-matrix(0,ncol=3,nrow=3)
                    diag(cov9)<-sqrt(c(diag(coefs[[2]])[c(20)
],0,diag(coefs[[2]])[c(21)]))
                    cov9[2,2]<-sqrt(mean(cov1[2,2],cov2[2,2],
cov3[2,2]))
                    cov9<-cov9%*%cormat%*%cov9
```

```
                    S<-list(cov1,cov2,cov3,cov4,cov5,cov6,cov7
,cov8,cov9)
                    S1<-S[1:3]
                    S2<-list(cov1,cov2,cov3,cov4[c(1,2),c(1,2)
],cov5[c(1,2),c(1,2)],cov6[c(1,2),c(1,2)],cov7[c(1,3),c(1,3)],
cov8[c(1,3),c(1,3)],cov9[c(1,3),c(1,3)])
                    Sigma<-as.matrix(bdiag(S2))

                    beta_t<-try(coef(mvmeta(formula=cbind(data
[,1],data[,2],data[,3])~1,S=S,method="fixed")),TRUE)
                    beta_t1<-try(coef(mvmeta(formula=cbind(
data1[,1],data1[,2],data1[,3])~1,S=S1,method="fixed")),TRUE)


                    if(inherits(beta_t,"try-error")==TRUE |
inherits(beta_t1,"try-error")==TRUE){

                            message(paste(i,",",j,"."),
appendLF=FALSE)
                            i<-i

                    }else{

                            W2<-cbind(W,coef=coefs[[1]])
                            result_1<-try(nlm(resfun,beta_t1)$
estimate,TRUE)



                            if(inherits(result_1,"try-error")
==TRUE ){
                                    bias0[i,j]<-NA
                                    bias1[i,j]<-NA
                                    bias2[i,j]<-NA

                                    beta_t00[i,j]<-NA
                                    beta_t01[i,j]<-NA
                                    beta_t02[i,j]<-NA

                                    beta_t10[i,j]<-NA
                                    beta_t11[i,j]<-NA
                                    beta_t12[i,j]<-NA

                                    i<-i+1
                            }else{

                                    bias0[i,j]<-1-result_1[1]
```

```
                                        bias1[i,j]<-beta1[j]-
    result_1[2]
                                        bias2[i,j]<-1-result_1[3]

                                        beta_t00[i,j]<-1-beta_t[1]
                                        beta_t01[i,j]<-beta1[j]-
    beta_t[2]
                                        beta_t02[i,j]<-1-beta_t[3]

                                        beta_t10[i,j]<-1-beta_t1
    [1]
                                        beta_t11[i,j]<-beta1[j]-
    beta_t1[2]
                                        beta_t12[i,j]<-1-beta_t1
    [3]

                                        i<-i+1
                                }
                        }


                }else{

                        i<-i
                }

        }
}


r1<-apply(bias0,2,function(x) mean(x,na.rm=T))
r2<-apply(bias1,2,function(x) mean(x,na.rm=T))
r3<-apply(bias2,2,function(x) mean(x,na.rm=T))

r4<-apply(beta_t00,2,function(x) mean(x,na.rm=T))
r5<-apply(beta_t01,2,function(x) mean(x,na.rm=T))
r6<-apply(beta_t02,2,function(x) mean(x,na.rm=T))

r7<-apply(beta_t10,2,function(x) mean(x,na.rm=T))
r8<-apply(beta_t11,2,function(x) mean(x,na.rm=T))
r9<-apply(beta_t12,2,function(x) mean(x,na.rm=T))


rbind(r1,r2,r3,r7,r8,r9,r4,r5,r6)


m1<-apply(bias0,2,function(x) mean(x^2,na.rm=T))
m2<-apply(bias1,2,function(x) mean(x^2,na.rm=T))
```

```
m3<-apply(bias2,2,function(x) mean(x^2,na.rm=T))

m4<-apply(beta_t00,2,function(x) mean(x^2,na.rm=T))
m5<-apply(beta_t01,2,function(x) mean(x^2,na.rm=T))
m6<-apply(beta_t02,2,function(x) mean(x^2,na.rm=T))

m7<-apply(beta_t10,2,function(x) mean(x^2,na.rm=T))
m8<-apply(beta_t11,2,function(x) mean(x^2,na.rm=T))
m9<-apply(beta_t12,2,function(x) mean(x^2,na.rm=T))

rbind(m1,m2,m3,m7,m8,m9,m4,m5,m6)
```

Appendix B

The exponential family and the partition function

In general, define the probability density function $p(\boldsymbol{x}|\boldsymbol{\theta})$, for $\boldsymbol{x} = (x_1, \ldots, x_m) \in \chi^m$ and $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$, and it is said to be exponential family if as follow;

$$\begin{aligned} p(\boldsymbol{x}|\boldsymbol{\theta}) &= \frac{1}{Z(\boldsymbol{\theta})} h(\boldsymbol{x}) \exp(\boldsymbol{\theta}\phi(\boldsymbol{x})) \\ &= h(\boldsymbol{x}) \exp(\boldsymbol{\theta}\phi(\boldsymbol{x}) - A(\boldsymbol{\theta})) \\ &= h(\boldsymbol{x}) \exp(\eta(\boldsymbol{\theta}^T)\phi(\boldsymbol{x}) - A(\eta(\boldsymbol{\theta}))) \end{aligned}$$

where

$$\begin{aligned} Z(\boldsymbol{\theta}) &= \int_{\chi^m} h(\boldsymbol{x}) \exp(\boldsymbol{\theta}\phi(\boldsymbol{x})) \\ A(\boldsymbol{\theta}) &= \log Z(\boldsymbol{\theta}) \end{aligned}$$

Here we call; $\boldsymbol{\theta}$ is the natural parameter or the canonical parameter, $\phi(\boldsymbol{x})$ is the sufficient statistic, $Z(\boldsymbol{\theta})$ is the partition function, $A(\boldsymbol{\theta})$ is the log partition function or the cumulant function, $h(\boldsymbol{x})$ is the scaling constant, often $= 1$, and $\eta(\boldsymbol{\theta})$ is a mapping of $\boldsymbol{\theta}$ to the canonical parameters. In addition, I note the following;

- If $dim(\boldsymbol{\theta}) < dim(\eta(\boldsymbol{\theta}))$, it is called a curved exponential family, that means we have more sufficient statistics than parameters.
- If $dim(\boldsymbol{\theta}) = dim(\eta(\boldsymbol{\theta}))$, it is called a canonical form.
- If $\phi((x)) = \boldsymbol{x}$, it is called a natural exponential family.

An important property of the exponential family and the log partition function is that the log partition function can be used to derive the cumulants of the sufficient statistics. That is why $A(\boldsymbol{\theta})$ is called the cumulant function. The derivation is as follows;

$$\begin{aligned}
\frac{dA(\theta)}{d\theta} &= \frac{d}{d\theta}\left(\log \int \exp(\theta\phi(x))h(x)dx\right) \\
&= \frac{\int \phi(x)\exp(\theta\phi(x))h(x)dx}{\exp(A(\theta))} \\
&= \int \phi(x)\exp(\theta\phi(x) - A(\theta))h(x)dx \\
&= \int \phi(x)p(x)dx \\
&= E[\phi(x)] = \text{Expectation of the sufficient statistics}
\end{aligned}$$

$$\begin{aligned}
\frac{d^2A(\theta)}{d\theta^2} &= \int \phi(x)\exp(\theta\phi(x) - A(\theta))h(x)(\phi(x) - A^{'}(\theta))dx \\
&= \int \phi(x)p(x)(\phi(x) - A^{'}(\theta))dx \\
&= \int \phi^2(x)p(x)dx - A^{'}(\theta)\int \phi(x)p(x)dx \\
&= E[\phi^2(x)] - E[\phi(x)]^2 \\
&\quad (\because A^{'}(\theta) = \frac{dA}{d\theta} = E[\phi(x)]) \\
&= Var[\phi(x)] = \text{Variance of the sufficient statistics}
\end{aligned}$$

More detailed explanation can be found in elsewhere such as [100, 77, 76]

Appendix C

Derivation of omitted variable bias formula

Suppose $\boldsymbol{X}$ and $\boldsymbol{Z}$ follow multivariate normal distribution, $N\left(\begin{pmatrix} \boldsymbol{\mu_X^T} \\ \boldsymbol{\mu_Z^T} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XZ} \\ \boldsymbol{\Sigma}_{ZX} & \boldsymbol{\Sigma}_{ZZ} \end{pmatrix}\right)$, and the distribution of $\boldsymbol{Z}$ conditional on $\boldsymbol{X}$ can be denoted as $\boldsymbol{Z}|\boldsymbol{X} \sim N(\boldsymbol{\mu_Z} + \boldsymbol{\Sigma_{ZX}}\boldsymbol{\Sigma_{XX}^{-1}}(\boldsymbol{X} - \boldsymbol{\mu_X}), \boldsymbol{\Sigma}_{XX} - \boldsymbol{\Sigma}_{XZ}\boldsymbol{\Sigma}_{ZZ}^{-1}\boldsymbol{\Sigma}_{ZX})$ [101]. Therefore, the conditional expectation of $\boldsymbol{Z}$ can be expressed as $\boldsymbol{\Gamma}_0 + \boldsymbol{X}\boldsymbol{\Gamma}_1$, where $\boldsymbol{\Gamma}_0 = \boldsymbol{\mu_Z} - \boldsymbol{\Sigma_{ZX}}\boldsymbol{\Sigma_{XX}^{-1}}\boldsymbol{\mu_X}$ and $\boldsymbol{\Gamma}_1 = (\boldsymbol{\Sigma_{ZX}}\boldsymbol{\Sigma_{XX}^{-1}})^T$.

Then (5) becomes

$$\begin{aligned} E\left[\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\gamma}^*)\right] &= E\left[\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{Z}\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\gamma}^*)\right] \\ &= E\left[\boldsymbol{X}^T\{\boldsymbol{X}\boldsymbol{\alpha} + (\boldsymbol{\Gamma}_0 + \boldsymbol{X}\boldsymbol{\Gamma}_1)\boldsymbol{\beta} - \boldsymbol{X}\boldsymbol{\gamma}^*\}\right] \\ &= \int \left\{\boldsymbol{X}^T\boldsymbol{X}(\boldsymbol{\alpha} + \boldsymbol{\Gamma_1}\boldsymbol{\beta} - \boldsymbol{\gamma}^*) + \boldsymbol{X}^T\boldsymbol{\Gamma}_0\boldsymbol{\beta}\right\} p_{X_1,\ldots,X_m} dX_1 \ldots dX_m = \mathbf{0}, \end{aligned}$$

where $p_{X_1,\ldots,X_m}$ indicates the joint distribution of $X_1, \ldots, X_m$.

When $\boldsymbol{\mu_X} = \mathbf{0}$ and $\boldsymbol{\mu_Z} = \mathbf{0}$, this reduced to $\boldsymbol{\Sigma}_{XX}\left\{\boldsymbol{\alpha} + (\boldsymbol{\Sigma_{ZX}}\boldsymbol{\Sigma_{XX}^{-1}})^T\boldsymbol{\beta} - \boldsymbol{\gamma}^*\right\} = \mathbf{0}$. Then, finally we get

$$\boldsymbol{\gamma}^* = \boldsymbol{\alpha} + E[(\boldsymbol{X^T X})^{-1}\boldsymbol{X}^T\boldsymbol{Z}]\boldsymbol{\beta},$$

which is correspond to the result of Equation (2.17) in the main text.

Appendix D

Proof of the formula (2.22)

In general, we assume the following multiple regression model; $Y_i = \alpha_0 + \alpha_1 X_{i1} + \cdots + \alpha_s X_{is} + u_i$, and also assume

$$\underset{n\times s}{\bar{\boldsymbol{X}}} = \begin{pmatrix} \underset{n\times 1}{\bar{\boldsymbol{x}}_1} & \underset{n\times(s-1)}{\bar{\boldsymbol{Z}}} \end{pmatrix},$$

where $\bar{\boldsymbol{x}}_1$ is a deviation vector of X_{i1} from the average and

$$\underset{n\times(s-1)}{\bar{\boldsymbol{Z}}} = \begin{pmatrix} X_{12} - \bar{X}_2 & \cdots & X_{1s} - \bar{X}_s \\ \vdots & \ddots & \vdots \\ X_{n2} - \bar{X}_2 & \cdots & X_{ns} - \bar{X}_s \end{pmatrix}.$$

Let us denote

$$(\bar{\boldsymbol{X}}^T \bar{\boldsymbol{X}})^{-1} = \begin{pmatrix} \bar{\boldsymbol{x}}_1{}^T \bar{\boldsymbol{x}}_1 & \bar{\boldsymbol{x}}_1^T \bar{\boldsymbol{Z}} \\ \bar{\boldsymbol{Z}}^T \bar{\boldsymbol{x}}_1 & \bar{\boldsymbol{Z}}^T \bar{\boldsymbol{Z}} \end{pmatrix}^{-1} = \begin{pmatrix} B_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix},$$

thus we obtain $\mathrm{Var}(\hat{\alpha}_1) = \hat{\sigma}^2 B_{11}$. From the matrix inversion lemma, the following equation can be calculated;

$$\begin{aligned} B_{11} &= \left(\bar{\boldsymbol{x}}_1^T \bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_1^T \bar{\boldsymbol{Z}} (\bar{\boldsymbol{Z}}^T \bar{\boldsymbol{Z}})^{-1} \bar{\boldsymbol{Z}}^T \bar{\boldsymbol{x}}_1 \right)^{-1} \\ &= (\bar{\boldsymbol{x}}_1{}^T \bar{\boldsymbol{x}}_1)^{-1} \left(1 - \frac{\bar{\boldsymbol{x}}_1^T \bar{\boldsymbol{Z}} (\bar{\boldsymbol{Z}}^T \bar{\boldsymbol{Z}})^{-1} \bar{\boldsymbol{Z}}^T \bar{\boldsymbol{x}}_1}{\bar{\boldsymbol{x}}_1^T \bar{\boldsymbol{x}}_1} \right)^{-1}, \end{aligned}$$

where $(\bar{\boldsymbol{Z}}^T\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}^T\bar{\boldsymbol{x}}_1$ can be regarded as the estimates $\hat{\boldsymbol{\beta}}$ of coefficients $\boldsymbol{\beta}$ in the regression model $\bar{\boldsymbol{x}}_1 = \bar{\boldsymbol{Z}}\boldsymbol{\beta} + \boldsymbol{e}$. Therefore, $\dfrac{\bar{\boldsymbol{x}}_1^T\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}^T\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}^T\bar{\boldsymbol{x}}_1}{\bar{\boldsymbol{x}}_1^T\bar{\boldsymbol{x}}_1} = \dfrac{\bar{\boldsymbol{x}}_1^T\bar{\boldsymbol{Z}}\hat{\boldsymbol{\beta}}}{\bar{\boldsymbol{x}}_1^T\bar{\boldsymbol{x}}_1}$ describes the proportion of variability that is covered by the regression compared with the total variability of $\bar{\boldsymbol{x}}_1$ and this is same as the definition of a coefficient of determination.

Thus, we can obtain

$$B_{11} = \frac{1}{n\mathrm{Var}(X_1)(1 - R_1^2)},$$

where R_1^2 indicates the coefficient of determination of regression of $\bar{\boldsymbol{x}}_1$ on other variables $\bar{\boldsymbol{Z}}$ and this is exactly same with the coefficient of determination of regression of X_1 on other variables $X_2, \ldots, X_s$.

Bibliography

[1] Cochrane. Glossary from cocnrahe community, July 2015. URL http://community.cochrane.org/glossary.

[2] Oxford centre for evidence-based medicine: Levels of evidence, July 2015. URL http://www.cebm.net/category/ebm-resources/.

[3] Gene V Glass. Primary, secondary, and meta-analysis of research. *Educational researcher*, pages 3–8, 1976.

[4] Ryan T Williams. *Using robust standard errors to combine multiple regression estimates with meta-analysis*. PhD thesis, Loyola University Chicago, 2012.

[5] Alexander J Sutton and Julian Higgins. Recent developments in meta-analysis. *Statistics in Medicine*, 27(5):625–650, 2008.

[6] Larry V Hedges and Ingram Olkin. *Statistical method for meta-analysis*. Academic press, 1985.

[7] Anne Whitehead. *Meta-analysis of controlled clinical trials*, volume 7. John Wiley Sons, 2002.

[8] Larry V Hedges and Therese D Pigott. The power of statistical tests in meta-analysis. *Psychological methods*, 6(3):203, 2001.

[9] Julian Higgins, Simon G Thompson, and David J Spiegelhalter. A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(1):137–159, 2009.

[10] Rebecca DerSimonian and Nan Laird. Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3):177–188, 1986.

[11] Sharon-Lise T Normand. Tutorial in biostatistics meta-analysis: formulating, evaluating, combining, and reporting. *Statistics in Medicine*, 18(3): 321–359, 1999.

[12] Eugene Demidenko. *Mixed models: theory and applications with R*. John Wiley Sons, 2013.

[13] Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

[14] Teresa C Smith, David J Spiegelhalter, and Andrew Thomas. Bayesian approaches to random-effects meta-analysis: a comparative study. *Statistics in Medicine*, 14(24):2685–2699, 1995.

[15] CN Morris and SL Normand. Hierarchical models for combining information and for meta-analyses. *Bayesian statistics*, 4:321–344, 1992.

[16] Richard D Riley, Keith R Abrams, Alexander J Sutton, Paul C Lambert, and John R Thompson. Bivariate random-effects meta-analysis and the estimation of between-study correlation. *BMC Medical Research Methodology*, 7(1):3, 2007.

[17] Yong Chen, Chuan Hong, and Richard D Riley. An alternative pseudolikelihood method for multivariate random-effects meta-analysis. *Statistics in Medicine*, 34(3):361–380, 2015.

[18] Hans C Van Houwelingen, Lidia R Arends, and Theo Stijnen. Tutorial in biostatistics. advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, 21(4):589–624, 2002.

[19] Dan Jackson, Richard Riley, and Ian R White. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*, 30(20):2481–2498, 2011.

[20] Richard D Riley, John R Thompson, and Keith R Abrams. An alternative model for bivariate random-effects meta-analysis when the within-study correlations are unknown. *Biostatistics*, 9(1):172–186, 2008.

[21] Julian Higgins, Simon Thompson, Jonathan Deeks, and Douglas Altman. Statistical heterogeneity in systematic reviews of clinical trials: a critical

appraisal of guidelines and practice. *Journal of health services research policy*, 7(1):51–61, 2002.

[22] Bertrand Baujat, Cédric Mahé, Jean-Pierre Pignon, and Catherine Hill. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Statistics in Medicine*, 21(18):2641–2652, 2002.

[23] PP Glasziou and SL Sanders. Investigating causes of heterogeneity in systematic reviews. *Statistics in Medicine*, 21(11):1503–1511, 2002.

[24] Fujian Song. Exploring heterogeneity in meta-analysis: is the l'abbe plot useful? *Journal of clinical epidemiology*, 52(8):725–730, 1999.

[25] Tania B Huedo-Medina, Julio Sánchez-Meca, Fulgencio Marín-Martínez, and Juan Botella. Assessing heterogeneity in meta-analysis: Q statistic or i^2 index? *Psychological methods*, 11(2):193, 2006.

[26] William G Cochran. The $\chi 2$ test of goodness of fit. *The Annals of Mathematical Statistics*, pages 315–345, 1952.

[27] William G Cochran. The combination of estimates from different experiments. *Biometrics*, 10(1):101–129, 1954.

[28] Dan Jackson. The power of the standard test for the presence of heterogeneity in meta-analysis. *Statistics in Medicine*, 25(15):2688–2699, 2006.

[29] Julian PT Higgins, Simon G Thompson, et al. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11):1539–1558, 2002.

[30] Patricia B Elmore and Paula L Woehlke. Twenty years of research methods employed in" american educational research journal,"" educational researcher," and" review of educational research.". *American Educational Research Journal*, 1998.

[31] Michael Matheny, Melissa L McPheeters, Allison Glasser, Nate Mercaldo, Rachel B Weaver, Rebecca N Jerome, Rachel Walden, J Nikki McKoy, Jason Pritchett, and Chris Tsai. Systematic review of cardiovascular disease

risk assessment tools. *Evidence Syntheses/Technology Assessments*, (85), 2011.

[32] Ali Abbasi, Linda M Peelen, Eva Corpeleijn, Yvonne T van der Schouw, Ronald P Stolk, Annemieke MW Spijkerman, Daphne L van der A, Karel GM Moons, Gerjan Navis, Stephan JL Bakker, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ: British Medical Journal*, 345, 2012.

[33] Betsy Jane Becker and Meng-Jia Wu. The synthesis of regression slopes in meta-analysis. *Statistical Science*, pages 414–429, 2007.

[34] Glenn A Walker and John G Saw. The distribution of linear combinations of t-variables. *Journal of the American Statistical Association*, 73(364): 876–878, 1978.

[35] Tom D Stanley and Stephen B Jarrell. Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys*, 3(2): 161–170, 1989.

[36] Richard R Lau, Lee Sigelman, Caroline Heldman, and Paul Babbitt. The effects of negative political advertisements: A meta-analytic assessment. *American Political Science Review*, 93(04):851–875, 1999.

[37] Neil H Timm. Estimating effect sizes in exploratory experimental studies when using a linear model. *The American Statistician*, 58(3), 2004.

[38] Sander Greenland. Quantitative methods in the review of epidemiologic literature. *Epidemiologic reviews*, 9(1):1–30, 1987.

[39] Sander Greenland and Matthew P Longnecker. Methods for trend estimation from summarized dose-response data, with applications to meta-analysis. *American journal of epidemiology*, 135(11):1301–1309, 1992.

[40] Jian Qing Shi and JB Copas. Meta-analysis for trend estimation. *Statistics in medicine*, 23(1):3–19, 2004.

[41] Yinghui Wei and Julian Higgins. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Statistics in medicine*, 32(7):1191–1205, 2013.

[42] CS Berkey, JJ Anderson, and DC Hoaglin. Multiple-outcome meta-analysis of clinical trials. *Statistics in Medicine*, 15(5):537–557, 1996.

[43] Richard D Riley. Multivariate meta-analysis: the effect of ignoring within-study correlation. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(4):789–811, 2009.

[44] Richard D Riley and Ewout W Steyerberg. Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods*, 1(1):2–19, 2010.

[45] MC Simmonds. *Statistical methodology of individual patient data*. PhD thesis, Ph. D. Thesis, 2005.

[46] Taye H Hamza, Hans C van Houwelingen, and Theo Stijnen. The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of clinical epidemiology*, 61(1):41–51, 2008.

[47] Thomas A Trikalinos, David C Hoaglin, and Christopher H Schmid. Empirical and simulation-based comparison of univariate and multivariate meta-analysis for binary outcomes. *Methods Research Reports*, 2013.

[48] Theo Stijnen, Taye H Hamza, and Pinar Özdemir. 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, 2010.

[49] Stephen W Raudenbush, Betsy J Becker, and Hripsime Kalaian. Modeling multivariate effect sizes. *Psychological Bulletin*, 103(1):111, 1988.

[50] Thomas Debray, Hendrik Koffijberg, Yvonne Vergouwe, Karel GM Moons, and Ewout W Steyerberg. Aggregating published prediction models with individual participant data: a comparison of different approaches. *Statistics in Medicine*, 31(23):2697–2712, 2012.

[51] Meng-Jia Wu and Betsy Jane Becker. Synthesizing regression results: a factored likelihood method. *Research Synthesis Methods*, 4(2):127–143, 2013.

[52] John M Neuhaus and Nicholas P Jewell. A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika*, 80 (4):807–815, 1993.

[53] Y Yi Grace and Nancy Reid. A note on mis-specified estimating functions. *Statistica Sinica*, 20:1749–1769, 2010.

[54] Jan Salomon Cramer. Omitted variables and misspecified disturbances in the logit model. Technical report, Tinbergen Institute Discussion Paper, 2005.

[55] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, pages 1–25, 1982.

[56] The Fibrinogen Studies Collaboration. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. *Statistics in Medicine*, 28(8):1218, 2009.

[57] David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

[58] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994.

[59] EW Steyerberg, MJ Eijkemans, JC Van Houwelingen, KL Lee, and JD Habbema. Prognostic models based on literature and individual patient data in logistic regression analysis. *Statistics in medicine*, 19(2):141–160, 2000.

[60] Ewout W Steyerberg, Marinus JC Eijkemans, Frank E Harrell, and J Dik F Habbema. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Statistics in medicine*, 19(8):1059–1079, 2000.

[61] Matthieu Resche-Rigon, Ian R White, Jonathan W Bartlett, Sanne AE Peters, and Simon G Thompson. Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Statistics in Medicine*, 32(28):4890–4905, 2013.

[62] Donald B Rubin. *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley Sons, 2004.

[63] Joseph L Schafer. Multiple imputation: a primer. *Statistical methods in medical research*, 8(1):3–15, 1999.

[64] Geert Molenberghs, Herbert Thijs, Ivy Jansen, Caroline Beunckens, Michael G Kenward, Craig Mallinckrodt, and Raymond J Carroll. Analyzing incomplete longitudinal clinical trial data. *Biostatistics*, 5(3):445–464, 2004.

[65] Joseph L Schafer and Recai M Yucel. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of computational and Graphical Statistics*, 11(2):437–457, 2002.

[66] Joop Hox and J Kyle Roberts. *Handbook of advanced multilevel analysis*. Psychology Press, 2011.

[67] Ian R White, Patrick Royston, and Angela M Wood. Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in medicine*, 30(4):377–399, 2011.

[68] Karel GM Moons, Rogier ART Donders, Theo Stijnen, and Frank E Harrell. Using the outcome for imputation of missing predictor values was preferred. *Journal of clinical epidemiology*, 59(10):1092–1101, 2006.

[69] Vidyadhar P Godambe. An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, pages 1208–1211, 1960.

[70] Peter McCullagh and Robert Tibshirani. A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 325–344, 1990.

[71] T Yanagimoto and E Yamamoto. *The role of unbiasedness in estimating equations*. 1991.

[72] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

[73] Peter J Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–233, 1967.

[74] AM Walker. On the asymptotic behaviour of posterior distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 80–88, 1969.

[75] John A Nelder and RJ Baker. Generalized linear models. *Encyclopedia of Statistical Sciences*, 1972.

[76] Peter McCullagh and John A Nelder. *Generalized linear models*, volume 37. CRC press, 1989.

[77] William H Greene. *Econometric analysis*. Pearson Education India, 2003.

[78] Dungang Liu, Regina Liu, and Minge Xie. Multivariate meta-analysis of heterogeneous studies using only summary statistics: efficiency and robustness. *Journal of the American Statistical Association*, (just-accepted): 00–00, 2014.

[79] Wei-Hsiung Chao, Mari Palta, and Terry Young. Effect of omitted confounders on the analysis of correlated binary data. *Biometrics*, pages 678–689, 1997.

[80] Norman L Johnson and Samuel Kotz. *Distributions in Statistics: Continuous Univariate Distributions: Vol.: 2*. Houghton Mifflin, 1970.

[81] Jan Salomon Cramer. *Logit models from economics and other fields*. Cambridge University Press, 2003.

[82] Richard D Riley, KR Abrams, PC Lambert, AJ Sutton, and JR Thompson. An evaluation of bivariate random-effects meta-analysis for the joint

synthesis of two correlated outcomes. *Statistics in Medicine*, 26(1):78–97, 2007.

[83] H Van Houwelingen, L Arends, and T Stijnen. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medicine*, 21:589–624, 2002.

[84] Jamie J Kirkham, Richard D Riley, and Paula R Williamson. A multivariate meta-analysis approach for reducing the impact of outcome reporting bias in systematic reviews. *Statistics in Medicine*, 31(20):2179–2195, 2012.

[85] In-Sun Nam, Kerrie Mengersen, and Paul Garthwaite. Multivariate meta-analysis. *Statistics in Medicine*, 22(14):2309–2333, 2003.

[86] Alan S Go, Dariush Mozaffarian, Veronique L Roger, Emelia J Benjamin, Jarett D Berry, Michael J Blaha, Shifan Dai, Earl S Ford, Caroline S Fox, Sheila Franco, et al. Heart disease and stroke statistics–2014 update: a report from the american heart association. *Circulation*, 129(3):e28, 2014.

[87] KC Johnston, AF Connors, DP Wagner, WA Knaus, X-Q Wang, E Clarke Haley, et al. A predictive risk model for outcomes of ischemic stroke. *Stroke*, 31(2):448–455, 2000.

[88] Shaw Watanabe, Shoichiro Tsugane, Tomotaka Sobue, Masamitsu Konishi, and Shunroku Baba. Study design and organization of the jphc study. *Journal of epidemiology*, 11(6sup):3–7, 2001.

[89] Shoichiro Tsugane and Norie Sawada. The jphc study: design and some findings on the typical japanese diet. *Japanese journal of clinical oncology*, 44(9):777–782, 2014.

[90] AE Walker, M Robins, and FD Weinfeld. The national survey of stroke. clinical findings. *Stroke; a journal of cerebral circulation*, 12(2 Pt 2 Suppl 1):I13, 1981.

[91] Hiroyasu Iso, Kathryn Rexrode, Charles H Hennekens, and Joann E Manson. Application of computer tomography-oriented criteria for stroke subtype classification in a prospective study. *Annals of epidemiology*, 10(2): 81–87, 2000.

[92] Hiroyuki Noda, Hiroyasu Iso, Isao Saito, Masamitsu Konishi, Manami Inoue, and Shoichiro Tsugane. The impact of the metabolic syndrome and its components on the incidence of ischemic heart disease and stroke: the japan public health center-based study. *Hypertension Research*, 32(4):289–298, 2009.

[93] I Saito, H Iso, Y Kokubo, M Inoue, and S Tsugane. Body mass index, weight change and risk of stroke and stroke subtypes: the japan public health center-based prospective (jphc) study. *International Journal of Obesity*, 35(2):283–291, 2010.

[94] Hiroshi Yatsuya, Hiroyasu Iso, Kazumasa Yamagishi, Yoshihiro Kokubo, Isao Saito, Kazuo Suzuki, Norie Sawada, Manami Inoue, and Shoichiro Tsugane. Development of a point-based prediction model for the incidence of total stroke japan public health center study. *Stroke*, 44(5):1295–1302, 2013.

[95] Margaret Sullivan Pepe. *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press, 2003.

[96] Gary S Collins, Joris A de Groot, Susan Dutton, Omar Omar, Milensu Shanyinde, Abdelouahid Tajar, Merryn Voysey, Rose Wharton, Ly-Mee Yu, Karel G Moons, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Medical Research Methodology*, 14(1):40, 2014.

[97] Rajat Raina, Andrew Y Ng, and Daphne Koller. Transfer learning by constructing informative priors. *Inductive transfer*, 10, 2006.

[98] Rich Caruana. A dozen tricks with multitask learning. In *Neural networks: tricks of the trade*, pages 165–191. Springer, 1998.

[99] Sebastian Thrun and Lorien Pratt. Learning to learn, chapter learning to learn: Introduction and overview. *Thrun and Pratt [1997a]*, pages 3–13, 1997.

[100] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[101] Shayle R Searle. *Matrix algebra useful for statistics*. 1982.