Bayesian Modeling for Environmental Association and Gene-Environment Interaction Under Complex Epidemiologic Study Designs

by

Shi Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2013

Doctoral Committee:

Professor Bhramar Mukherjee, Chair
Professor Stuart A. Batterman
Assistant Professor Veronica Berrocal
Assistant Professor Xiaoquan William Wen

To my family and my love...

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my advisor, Dr. Bhramar Mukherjee, for her insightful advice and kind help during my doctoral study in the University of Michigan. I would like to thank my committee member Dr. Stuart Batterman who has provided me valuable opportunities and experiences of research work. I extend my appreciation to my committee members Dr. Veronica Berrocal and Dr. Xiaoquan William Wen for serving in my committee. I would like to thank Dr. Brisa Sanchez for her encouraging advice and kind help for my master's course work.

I would like to thank Dr. Jeremy M.G. Taylor and Dr. Michael Boehnke for their insightful suggestions on the gene-environment interaction projects. I thank Dr. Heather M. Stringham and Dr. Michael Boehnke for sharing the genotyped data sets for investigating type 2 diabetes, and the D2D2007, DIAGEN, DPS, FUSION FSIGT, FUSION S2, HUNT, METSIM and TROMSO investigators for providing access to their data. I thank Dr. Malay Ghosh for his contribution on the case-crossover projects. I appreciate the help from Feng-Chiao Su, Elizabeth Wasilevich, Erika Garcia, Huda Elasaad and Robert Wahl for the Medicaid data analysis in the Detroit Asthma Morbidity, Air Quality and Traffic study. I would like to acknowledge the support of my dissertation projects by grants NSF DMS 1007494, NIH ES 20811 and CA 156608, EPA G2007 STAR A1, and DK 062370.

I thank, all of my teachers, mentors and supervisors from elementary to graduate school, sincerely. I thank, all of my friends, nearby or oceans apart.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

List of Abbreviations

- Chapter 2
  - conditional logistic regression (CLR)
  - polychotomous CLR (PCLR)
  - Markov chain Monte Carlo (MCMC)
  - maximum likelihood estimators (MLE)
  - iteratively re-weighted least squares (IRLS)
  - mean squared error (MSE)
  - polychotomous-category model (PCM)
  - adjacent-category model (ACM)
  - annual average daily traffic (AADT)
  - Akaike information criterion (AIC)
- Chapter 3
  - symmetric bidirectional designs (SBD)
  - time-stratified design (TSD)
- Chapter 4
  - genome-wide association studies (GWAS)
  - type 2 diabetes (T2D)
  - gene-environment interaction (GEI)
  - body mass index (BMI)
  - fixed-effects model (FEM)
  - random-effects model (REM)

- individual patient data (IPD)
- multivariate inverse-variance weighted (MIVW)
- univariate inverse-variance weighted (UIVW)
- meta-regression (MR)
- adaptively weighted estimator (AWE)
- asymptotic relative efficiency (ARE)
- minor allele frequency (MAF)
- Single Nucleotide Polymorphism (SNP)
- high-density lipoprotein (HDL)
- weighted least square (WLS)
- maximum likelihood estimate (MLE)
- Hardy-Weinberg equilibrium (HWE)

■ Chapter 5

- empirical Bayes (EB)
- partially empirical Bayes (PEB)
- inverse-variance weighted (IVW)
- hybrid inverse-variance weighted (HIVW)

CHAPTER I

Introduction

There has been substantial interest in the joint contribution of genetic (G) and environmental (E) factors to disease etiology, especially for complex human diseases. The definition of 'environment' can be quite broad, including demographic factors (age, gender), behavioral factors (smoking, alcohol consumption, medication use) and external factors (exposure to air pollution, radio-active substances). Different study designs can be gainfully employed depending on the nature of environmental exposure. This dissertation evolves around the theme of characterizing effect of environmental exposure on health outcomes under complex sampling designs in the first two projects. In the latter two, we consider the problem of meta-analysis of G-E interactions and how G-E independence and environmental heterogeneity across studies could influence the operating characteristics of several meta-analysis approaches. Thus, the dissertation makes important contributions to environmental epidemiology and its intersection with genetic epidemiology.

In the first project, we considered distance-odds models to investigate elevated disease odds around point sources of exposure, where there are sub-types within cases under a matched case-control design. We consider models analogous to the polychotomous logit models and adjacent-category logit models for categorical outcomes and extend them to the non-linear distance-odds context. Different inference methods including maxi-

mum likelihood, profile likelihood, iteratively re-weighted least squares and a hierarchical Bayesian approach using Markov chain Monte Carlo techniques were evaluated under these distance-odds models. We compare these methods using an extensive simulation study with multiple outcome categories and a non-linear distance-odds model. Bayesian methods appear to have advantages in terms of estimation stability, precision and interpretation over frequentist alternatives we considered. The proposed methods were applied to a population-based matched case-control study investigating associations between acute asthma outcomes and proximity of residence to major roads by analyzing Medicaid claims data for the pediatric asthma population in Detroit, MI, from 2004-2006, as part of the 'Detroit Asthma Morbidity, Air Quality and Traffic' (DAMAT) study.

The second project considered Bayesian analysis of time-series data under case-crossover designs. Case-crossover designs are widely used to study short-term exposure effects on the risk of acute adverse health events. While the frequentist literature on this topic is vast, there is no Bayesian work in this general area. The contribution of this project is two-fold. First, we establish Bayesian equivalence results that require characterization of the set of priors under which the posterior distributions of the risk ratio parameters based on a case-crossover and time-series analysis are identical. Second, we study more general inferential issues under case-crossover designs in a Bayesian framework. Traditionally, a conditional logistic regression is used for inference on risk-ratio parameters in case-crossover studies. We consider instead a more general full likelihood-based approach which makes less restrictive assumptions on the health risk models and exposure series. Formulation of a full likelihood leads to growth in the number of parameters proportional to the sample size and consequently maximum likelihood estimates are not consistent. We propose a semi-parametric Bayesian approach using a Dirichlet process prior to handle the random nuisance parameters that appear in a full likelihood formulation. We carry out a simula-

tion study to compare the Bayesian methods based on full and conditional likelihood with standard frequentist approaches for case-crossover and time-series analysis. The proposed methods are also illustrated through the DAMAT study, but instead of the distance to the major roads, we focus on the effect of ambient air pollutant concentrations on the acute asthma risk.

The third project considered meta-analysis of G-E interaction for quantitative traits. With heterogeneity in environmental covariate distributions across cohorts and obvious challenges with data harmonization involving various data sources, meta-analysis of studies of G-E interaction can often involve subtle statistical issues. In this project we study the effect of environmental covariate heterogeneity (within and between cohorts) on two approaches for fixed-effects meta-analysis: the standard inverse variance weighted meta-analysis and a meta-regression approach. Both are easily implemented for large scale association studies using summary statistics or published results. Though meta-regression is bias-prone and lacks power, the advantage is that the estimates corresponding to marginal genetic association analysis could be regressed on study-level environmental covariate means to screen for G-E interaction. Akin to the results obtained in Simmonds and Higgins (2007) in the context of detecting treatment-covariate interactions for randomized clinical trials, we obtain analytical efficiency/power expressions for both methods under the assumption of G-E independence. The relative efficiency/power of the two methods depend on the ratio of within versus between cohort variance of the environmental covariate.

In this project, instead of discretely choosing meta-analysis versus meta-regression based on this ratio as prescribed in Simmonds and Higgins (2007), or collecting multivariate summary statistics, we propose to use an adaptive combination of meta-analysis and meta-regression estimates that can be used as a default choice, retaining full efficiency

of the 'gold standard' pooled analysis for the interaction parameter using individual patient data (IPD) under certain natural assumptions. Lin and Zeng (2010) showed that a multivariate inverse-variance weighted estimator is asymptotically equivalent to the IPD estimator, given that all the common parameters with full information matrix under the fixed-effects model are pooled across all studies. They also characterized and quantified the efficiency loss of using an univariate (as a proper subset of the common parameters) inverse-variance estimator. We showed connection of our work to Lin and Zeng (2010). Essentially, the adaptive estimator gains efficiency by combining both meta-analysis versus meta-regression and using only univariate summary statistics from each study: estimates of marginal genetic association, interaction, their standard errors, as well as the mean of the environmental covariate. As a result, the adaptive approach bypasses issues with sharing of individual data across studies without sacrificing efficiency. We study the performance of all the methods under several common scenarios: (1) departures from G-E independence; (2) heterogeneity in minor allele frequencies across cohorts; (3) lack of common set of confounders to adjust in each study; (4) misspecification of the genetic susceptibility model (dominant/co-dominant/additive); (5) non-linear interaction. The results were illustrated through meta-analysis of interaction between Single Nucleotide Polymorphisms (SNPs) on the FTO gene and body mass index on high-density lipoprotein cholesterol data from a large consortium (Finland-United States Investigation of Noninsulin-dependent diabetes mellitus (FUSION) genetics study) of Type 2 diabetes.

The last project extends the work of project 3 to dichotomous traits under case-control studies. Gaining efficiency in studies of G-E interaction by exploiting independence between G and E under case-control sampling has been noted in multiple papers (Piegorsch et al. (1994); Umbach and Weinberg (1997); Chatterjee and Carroll (2005)). However, methods that use G-E independence assumption might produce severely biased estimates

if the assumption is violated. Several studies have addressed this issue and proposed more robust strategies for testing G-E interaction (Mukherjee and Chatterjee (2008), Mukherjee et al. (2008); Li and Conti (2009); Murcray et al. (2009)). Mukherjee and Chatterjee (2008) proposed a solution to the bias versus efficiency dilemma, using a retrospective method that allows for uncertainty around the assumption of G-E independence. Mukherjee and Chatterjee (2008) used the estimate of the uncertainty parameter in an empirical Bayes (EB) fashion to obtain a shrinkage estimator that 'shrinks' the maximum likelihood estimates (MLEs) of disease odds ratio parameters under G-E dependence to those under G-E independence, and showed how the shrinkage factor depends on these MLEs and their corresponding variances. Further theoretical development regarding this shrinkage estimator is presented in Chen et al. (2009). It was noted that this EB estimate can optimally trade off between bias and efficiency and provide increased power compared to a standard case-control analysis, with superior control of type 1 error when compared to a case-only analysis. As the G-E interactions detected so far only have small to modest effects, there are increasing demands for large sample sizes and collaboration across different study sites in order to perform a pooled or meta-analysis with high confidence and power. However, there are no papers thus far to study the role of G-E independence in a meta-analysis setting where the assumption could vary within each study.

In this project, we consider possible extensions of EB shrinkage estimators for a multiple-study setting, which uses the retrospective likelihood as the basis for influence and leverages the G-E independence assumption in a data-adaptive way. To handle this multiple-study problem, we particularly consider strategies to obtain a shrinkage factor in the EB estimator that can borrow strength across studies, under both IPD analysis using individual level data and meta-analysis using study level summary statistics. The proposed shrinkage estimator provides optimal choices for weights corresponding to constrained

and unconstrained models by using information on G-E association parameters derived from multiple studies/cohorts. Our work showed that this novel estimator has better MSE properties than IVW estimator pooling study specific constrained, unconstrained or EB estimators. The results were illustrated through the FUSION consortium, which has 6 different case-control studies that were treated as independent contributors to the analysis. We conducted an IPD/meta-analysis of interactions between SNPs on the FTO gene and environmental factors on the type 2 diabetes.

In the following chapters 2-5, I described the four dissertation projects sequentially. More specific background and literature review for these projects appear in their corresponding chapters. To summarize, this dissertation work is expected to contribute to important analytical, methodological questions that have relevance and applications in genetic and environmental epidemiology.

CHAPTER II

Point source modeling of matched case-control data with multiple disease sub-types

2.1 Introduction

In case-control designs, matching is commonly implemented in order to avoid bias due to potential confounders. In an individually matched case-control study, effects of potential risk factors are typically ascertained through a conditional likelihood approach such as conditional logistic regression (CLR) (Breslow et al. 1978). Extension of CLR to situations with multiple sub-types of cases or controls has been made through polychotomous CLR (PCLR), which is more efficient than carrying out separate CLRs for subgroups (Liang and Stewart 1987). Liang and Stewart (1987), Becher and Jöckel (1990), and Becher (1991) applied PCLR models to matched case-control studies with two control groups, typically hospital and population controls. Thomas et al. (1986) and Dubin and Pasternack (1986) applied PCLR models to analyze multiple disease groups with one set of controls. Sinha et al. (2004) considered a Bayesian semiparametric model for analyzing matched case-control data with multiple disease states and missing exposure values. Mukherjee et al. (2007) considered cases having multiple disease states with a natural ordering in matched case-control studies. Mukherjee et al. (2009) proposed a methodology to fit stratified proportional odds models by amalgamating conditional likelihoods obtained from all possible binary collapsings of the ordinal scale.

Studies since 1990's (Diggle (1990); Lawson (1993); Diggle and Rowlingson (1994); Diggle et al. (1997)) have investigated elevated risk of respiratory diseases around putative point sources of environmental pollution. Diggle and Rowlingson (1994) extended the exponential decay function described in Diggle (1990) to a situation where the decay function will equal to 1 after a certain threshold from the source. Diggle et al. (1997) then considered another threshold effect where the decay function does not decrease up to a certain distance from the source. Diggle et al. (2000) described an extension to matched case-control designs of the parametric modeling framework in Diggle (1990) using a conditional likelihood approach. Asthma and chronic obstructive airways disease were associated with proximity of residence to major roads in East London. There has been an increasing interest in modeling disease risk in relation to point sources of pollution in a Bayesian framework (Wakefield and Morris (2001); Lawson et al. (2003); Congdon and Congdon (2003)). Wakefield and Morris (2001) described a Bayesian hierarchical modeling of disease risk around a point source, embedding models proposed by Diggle et al. (1997). Issues of the sensitivity to prior specification for this class of models were discussed. Dreassi et al. (2008) performed a sensitivity analysis to investigate how the specification of the distance-odds functions and the choice of prior distributions affect results under case-control studies (Dreassi et al. 2008). Rodrigues et al. (2010) provided a semi-parametric approach for point process modeling using generalized additive model, and illustrated the flexibility of this approach with applications in epidemiology and criminology. All of the above spatial environmental epidemiology studies considered only the standard binary case-control states.

The purpose of this project is to incorporate the distance-odds model around point sources into the analysis of matched case-control data with multiple disease or control states. We extend the idea of the polychotomous logit model and the adjacent-category

logit model from the standard categorical data literature (Agresti 2002) to the non-linear distance-odds model framework. The extensions with non-linear odds function lead to some unique observations specific to the distance odds model. Maximum likelihood, profile likelihood, iteratively re-weighted least squares (IRLS) and a hierarchical Bayesian approach using MCMC are evaluated under the proposed models. Inference methods and various types of point source models are compared using an extensive simulation study. Simulation studies that compare the frequentist properties (such as bias, mean squared error (MSE) and coverage probability) of the proposed methods and models are not available in the literature, not even for binary case-control states.

2.2 Methods

Diggle *et al.* (1990 and 1994) proposed the distance-odds model for characterizing elevated risk around putative point sources of environmental pollution in case-control studies. The model assumes that the odds of disease, $r(x)$ as a function of distance x from the point source, is proportional to the decay function $f(x)$, as given below:

$$\text{(2.1)} \quad \frac{P(Y=1|x)}{P(Y=0|x)} = \frac{p(x)}{1-p(x)} = r(x) = \rho f(x) \quad \text{and} \quad f(x) = 1 + \alpha \exp\left(-(x/\beta)^2\right),$$

where Y is the disease status ($Y = 1$: case; $Y = 0$: control), x is the distance from the point source, ρ is the background odds of disease in the case-control population. (For a case-control study that is embedded in a cohort study, ρ is typically given by $\rho = (q_1/q_2)\kappa$, where κ is the background odds of disease in the study cohort, q_1 and q_2 are the proportions of cases and controls sampled from the cohort respectively.) The parameters (α, β) in model (2.1) have a natural interpretation: α is proportional to the disease odds at the point source ($\alpha = [r(0)/\rho] - 1$); β measures the rate of decay with increasing distance from the point source, in the unit of distance x. $(\alpha, \beta) \in (-1, \infty) \times (0, \infty)$. Under this model setting, as $x \to \infty$, we have $f(x) \to 1$ and the risk function $p(x) = P(Y = 1|x) =$

$\rho f(x)/(1+\rho f(x)) \to \rho/(1+\rho)$, i.e. the background risk in the case-control population (Diggle et al. 2000). Note that, if $f(x) = \exp(\beta x)$ is chosen with $r(x) = \rho f(x)$ in model (2.1), then one would have that $\log(r(x)) = \log(\rho) + \beta x$, which becomes the usual logistic regression model that assumes a linear distance-odds relationship with log odds ratio β and intercept $\log(\rho)$. However, usually the odds of disease changes non-linearly with increasing distance from the point source, e.g., with increasing distance to an industrial park, the odds of asthma might decrease much faster within 0-200 meters than within 1000-1200 meters. Another possible disadvantage of the log-linear model is that for $\beta < 0$ (that implies increasing odds with decreasing distance), $r(x) \to 0$ and $p(x) \to 0$ as $x \to \infty$, but these do not converge to background odds or risk which would be a desirable property. For non-rare diseases such as asthma, the log-linear distance-odds model is questionable. These disadvantages of log-linear model lead us to focus on the non-linear distance-odds model (2.1) proposed by Diggle (1990).

As an extension to model (2.1), Diggle and Rowlingson (1994) assumed multiplicative risk factors for the combined effects of S point sources, and allowed for covariate adjustment via additional log-linear terms. In the presence of S point sources and W spatially referenced covariates $Z_w(x), w = 1, ..., W$, the resulting distance-odds model is

$$r(\mathbf{x}) = \rho f(\mathbf{x}) \quad \text{and} \quad f(\mathbf{x}) = \exp\Big(\sum_{w=1}^{W} \phi_w Z_w(\mathbf{x})\Big) \prod_{s=1}^{S} f_s(x_s), \tag{2.2}$$

where $\mathbf{x} = (x_1, ..., x_S)$, x_s and $f_s(x_s)$ are the distance and the decay function for the *s*-th point source respectively. Here each $f_s(x_s)$ takes the same functional form as in model (2.1), that is, $f_s(x_s) = 1 + \alpha_s \exp(-(x_s/\beta_s)^2)$.

For a $1{:}M$ matched case-control study with N matched pairs, the risk of disease for an individual at distance x in the *i*-th stratum can be expressed as Diggle et al. (2000)

$$P_i(Y = 1|x) = \frac{r_i(x)}{1 + r_i(x)} = \frac{\rho_i f(x)}{1 + \rho_i f(x)}, \quad i = 1, ..., N,$$

where the baseline odds ρ_i for the *i*-th stratum can potentially vary across matched pairs under the matched case-control design. The conditional likelihood, given the exposure at distance $\mathbf{x_i} = (x_{i1}, x_{i2}, ..., x_{i(M+1)})$ for the *i*-th stratum, that the case is at distance x_{i1} is

$$\begin{aligned} L_i(\alpha, \beta) &= P(Y_{i1} = 1, Y_{i2} = ... = Y_{i(M+1)} = 0 | Y_{i1} + Y_{i2} + ... + Y_{i(M+1)} = 1, \mathbf{x_i}) \\ &= \frac{\frac{\rho_i f(x_{i1})}{\prod_{j=1}^{M+1}(1+\rho_i f(x_{ij}))}}{\frac{\rho_i f(x_{i1})}{\prod_{j=1}^{M+1}(1+\rho_i f(x_{ij}))} + \frac{\rho_i f(x_{i2})}{\prod_{j=1}^{M+1}(1+\rho_i f(x_{ij}))} + \cdots + \frac{\rho_i f(x_{i(M+1)})}{\prod_{j=1}^{M+1}(1+\rho_i f(x_{ij}))}} = \frac{f(x_{i1})}{\sum_{j=1}^{M+1} f(x_{ij})}, \end{aligned} \tag{2.3}$$

where Y_{ij} and x_{ij} are the disease status and distance for the *j*-th individual in the *i*-th stratum respectively, $i = 1, ..., N; j = 1, ..., M+1$. The general form of the conditional likelihood is (2.3). For one point source binary model, $f(x)$ is as given in (2.1), where as for multiple point sources binary model (with possible covariate adjustment) $f(x)$ is as given in (2.2).

Denote the conditional likelihood by L, the corresponding log-likelihood by l ($l = \log(L) = \sum_{i=1}^{N} \log(L_i) = \sum_{i=1}^{N} l_i$), and the parameters to be estimated by $\boldsymbol{\theta}$. The maximum likelihood estimates (MLEs) of $\boldsymbol{\theta} = (\alpha, \beta)$ in the one point source binary outcome model can be obtained by maximizing the logarithm of the conditional likelihood

$$l(\alpha, \beta) = \sum_{i=1}^{N} \log\Big(\frac{f(x_{i1})}{\sum_{j=1}^{M+1} f(x_{ij})}\Big) = \sum_{i=1}^{N} \log\Big(\frac{1 + \alpha \exp\left(-(x_{i1}/\beta)^2\right)}{\sum_{j=1}^{M+1}[1 + \alpha \exp\left(-(x_{ij}/\beta)^2\right)]}\Big).$$

Similarly, the MLEs of $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}) = (\alpha_1, ..., \alpha_S, \beta_1, ..., \beta_S, \phi_1, ..., \phi_W)$ in the S point sources binary outcome model with W covariates can be obtained by maximizing

$$l(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{i=1}^{N} \log\Big(\frac{\exp\left(\sum_{w=1}^{W} \phi_w Z_w(\boldsymbol{x_{i1}})\right) \prod_{s=1}^{S}(1 + \alpha_s \exp\left(-(x_{i1s}/\beta_s)^2\right))}{\sum_{j=1}^{M+1}\left[\exp\left(\sum_{w=1}^{W} \phi_w Z_w(\boldsymbol{x_{ij}})\right) \prod_{s=1}^{S}(1 + \alpha_s \exp\left(-(x_{ijs}/\beta_s)^2\right))\right]}\Big),$$

where $\boldsymbol{x_{ij}} = (x_{ij1}, ..., x_{ijS})$ and x_{ijs} is the distance of the *j*-th individual in the *i*-th stratum from the *s*-th point source. More detailed discussion of parameter estimation and inference for the models with binary outcomes can be found in Diggle et al. (2000).

2.2.1 Distance-odds model with polychotomous outcome

In this section, the distance-odds model is extended to situations where cases can have multiple disease states. Without loss of generality, the methods and formulation in the following sections are illustrated for a $1{:}M$ matched case-control data set with N matched pairs, where outcomes can belong to one of the K disease categories (For example, with $K = 2$. poor prognosis: $Y = 2$; fair prognosis: $Y = 1$) and one control group ($Y = 0$). These methods can be readily applied to situations with multiple control states, and to situations with variable matching ratios. The distance-odds model is adapted to both polychotomous-category model (PCM) and adjacent-category model (ACM) setting. The PCMs are considered when one tries to distinguish nominal disease sub-types to the controls. The ACMs are more appropriate when there is a natural ordering of the disease sub-classifications.

Polychotomous-category distance-odds model

For the polychotomous-category model, the odds of disease for the *j*-th individual in the *i*-th stratum at distance x_{ij} is modeled as

(2.4)

$$r^k(x_{ij}) = \frac{P(Y_{ij} = k|x_{ij})}{P(Y_{ij} = 0|x_{ij})} = \rho_{ik} f_k(x_{ij}), \quad i = 1, ..., N; j = 1, ..., M + 1; k = 1, ..., K,$$

where the baseline odds ρ_{ik} can potentially vary across matched pairs i and disease categories k, and the distance-odds function $f_k(x)$ can also vary among disease categories. Note that, if $f_k(x) = \exp(\beta_k x)$ is chosen in model (2.4) with multiplicative nuisance parameters $\rho_{ik} = \gamma_i \times \lambda_k$, one would have that

(2.5)
$$\log(r^k(x_{ij})) = \log\left(\frac{P(Y_{ij} = k|x_{ij})}{P(Y_{ij} = 0|x_{ij})}\right) = \log(\gamma_i) + \log(\lambda_k) + \beta_k x_{ij},$$

which becomes the polychotomous logistic regression models (Agresti 2002) that assumes a linear distance-odds relationship. Non-linear distance-odds models such as (2.1) are desired, with advantages over log-linear models. Using the K equations in (2.4) along with one more constraint that $\sum_{k=0}^{K} P(Y_{ij} = k|x_{ij}) = 1$, the risk of disease can be written in terms of ρ_{ik} and f_k for the corresponding individual, i.e.

$$\begin{aligned} P(Y_{ij} = 0|x_{ij}) &= \frac{1}{1 + \sum_{k=1}^{K}[\rho_{ik} f_k(x_{ij})]}, \\ P(Y_{ij} = k|x_{ij}) &= \frac{\rho_{ik} f_k(x_{ij})}{1 + \sum_{k=1}^{K}[\rho_{ik} f_k(x_{ij})]}, \quad k = 1, ..., K. \end{aligned}$$

Let k_i denote the disease states of the case subject in matched set *i*, $k_i \in (1, ..., K)$. The conditional likelihood for the *i*-th stratum, given a matched case-control pair at distance $\mathbf{x_i} = (x_{i1}, x_{i2}, ..., x_{i(M+1)})$, that the case (in category k_i) is at distance x_{i1} is

$$\begin{aligned} L_i^{k_i} &= P(Y_{i1} = k_i, Y_{i2} = ... = Y_{i(M+1)} = 0|Y_{i1} + Y_{i2} + ... + Y_{i(M+1)} = k_i, \mathbf{x_i}) \\ &= \frac{\rho_{ik_i} f_{k_i}(x_{i1}) / \prod_{j=1}^{M+1}[1 + \sum_{k=1}^{K} \rho_{ik} f_k(x_{ij})]}{\sum_{j=1}^{M+1} \rho_{ik_i} f_{k_i}(x_{ij}) / \prod_{j=1}^{M+1}[1 + \sum_{k=1}^{K} \rho_{ik} f_k(x_{ij})]} \\ &= \frac{f_{k_i}(x_{i1})}{\sum_{j=1}^{M+1} f_{k_i}(x_{ij})}. \end{aligned} \tag{2.6}$$

The general form of the conditional likelihood is (2.6). For one point source PCM, $f_k(x)$ is given as $f_k(x) = 1 + \alpha_k \exp(-(x/\beta_k)^2)$; for multiple point sources PCM $f_k(x)$ is given as $f_k(\boldsymbol{x}) = \exp(\sum_{t=1}^{T} \phi_{kt} Z_{kt}(\boldsymbol{x})) \prod_{s=1}^{S} f_{ks}(x_s)$ where $f_{ks}(x_s) = 1 + \alpha_{ks} \exp(-(x_s/\beta_{ks})^2)$.

Adjacent-category distance-odds model

For the adjacent-category model setting, the adjacent odds of disease between category K versus $K - 1$ for the *j*-th individual in the *i*-th stratum can be modeled as

$$r^k(x_{ij}) = \frac{P(Y_{ij} = k|x_{ij})}{P(Y_{ij} = k - 1|x_{ij})} = \rho_{ik} f_k(x_{ij}), \quad i = 1, ..., N; j = 1, ..., M+1; k = 1, ..., K. \tag{2.7}$$

Again, the baseline odds ρ_{ik} can vary across matched pairs i and disease categories k, and the distance-odds function $f_k(x)$ can vary across disease categories. One point source ACM and multiple point sources ACM (with possible covariate adjustment) can be formulated similarly as PCM with different choices of f_k. For these non-linear settings, ACM can not be represented as a re-parameterization of PCM as in log-linear models. Thus, both ACM and PCM are needed for ordered and nominal disease sub-classifications respectively. Note that if $f_k(x) = \exp(\beta_k^* x)$ is chosen in model (2.7) with multiplicative nuisance parameters $\rho_{ik} = \gamma_i^* \times \lambda_k^*$, one would have that

$$\log(r^k(x_{ij})) = \log\big(\frac{P(Y_{ij}=k|x_{ij})}{P(Y_{ij}=k-1|x_{ij})}\big) = \log(\gamma_i^*) + \log(\lambda_k^*) + \beta_k^* x_{ij}, \tag{2.8}$$

which reduces to the polychotomous logistic regression models in adjacent category setting (Agresti 2002) that assumes a linear distance-odds relationship. The risk of disease can be represented in terms of ρ_{ik} and f_k as

$$\begin{aligned} P(Y_{ij}=0|x_{ij}) &= \frac{1}{1+\sum_{k=1}^{K}[\prod_{h=1}^{k}\rho_{ih}f_h(x_{ij})]}, \\ P(Y_{ij}=k|x_{ij}) &= \frac{\prod_{h=1}^{k}\rho_{ih}f_h(x_{ij})}{1+\sum_{k=1}^{K}[\prod_{h=1}^{k}\rho_{ih}f_h(x_{ij})]}, \quad k=1,...,K. \end{aligned}$$

It follows that the conditional likelihood for the *i*-th stratum is

$$\begin{aligned} L_i^{k_i} &= P(Y_{i1}=k_i, Y_{i2}=...=Y_{i(M+1)}=0|Y_{i1}+Y_{i2}+...+Y_{i(M+1)}=k_i, \mathbf{x_i}) \\ &= \frac{\prod_{h=1}^{k_i}\rho_{ih}f_h(x_{i1})/\prod_{j=1}^{M+1}\left[1+\sum_{k=1}^{K}[\prod_{h=1}^{k}\rho_{ih}f_h(x_{ij})]\right]}{\sum_{j=1}^{M+1}[\prod_{h=1}^{k_i}\rho_{ih}f_h(x_{ij})]/\prod_{j=1}^{M+1}\left[1+\sum_{k=1}^{K}[\prod_{h=1}^{k}\rho_{ih}f_h(x_{ij})]\right]} \\ &= \frac{\prod_{h=1}^{k_i}f_h(x_{i1})}{\sum_{j=1}^{M+1}[\prod_{h=1}^{k_i}f_h(x_{ij})]}. \end{aligned} \tag{2.9}$$

One special case of interest is the homogeneity of the adjacent odds ratios (homogeneous ACM) with one unit increase in distance across case categories,

$$\begin{aligned} &\frac{r^K(x+1)}{r^K(x)} = \frac{r^{K-1}(x+1)}{r^{K-1}(x)} = \cdots = \frac{r^1(x+1)}{r^1(x)}, \quad \forall x \quad \Leftrightarrow \\ &\alpha_1 = \alpha_2 = ... = \alpha_K \quad \text{and} \quad \beta_1 = \beta_2 = ... = \beta_K. \end{aligned} \tag{2.10}$$

2.2.2 Estimation and inference

Maximum likelihood approach

Without loss of generality, the first subject in each stratum is always considered as the case when deriving the likelihood and fitting the models, i.e., $Y_{i1} = k_i, k_i \in (1, ..., K)$. Thus, the actual contribution of the *i*-th stratum to the conditional likelihood is $L_i^{k_i}$ as given in (2.6) for PCM or as given in (2.9) for ACM respectively. For example, the MLEs for ACM can be obtained by maximizing the logarithm of the conditional likelihood

$$\sum_{i=1}^{N} \log\left(L_i^{k_i}(\boldsymbol{\alpha}, \boldsymbol{\beta})\right) = \sum_{i=1}^{N} \log\Big(\frac{\prod_{h=1}^{k_i}(1 + \alpha_h \exp\left(-(x_{i1}/\beta_h)^2\right))}{\sum_{j=1}^{M+1}\prod_{h=1}^{k_i}(1 + \alpha_h \exp\left(-(x_{ij}/\beta_h)^2\right))}\Big), \tag{2.11}$$

or the following in the most general case with multiple sources and covariate adjustment

$$\sum_{i=1}^{N} \log\Big(\frac{\prod_{h=1}^{k_i}[\exp\left(\sum_{w=1}^{W}\phi_{hw}Z_{hw}(\boldsymbol{x_{i1}})\right)\prod_{s=1}^{S}(1 + \alpha_{hs} \exp\left(-(x_{i1s}/\beta_{hs})^2\right))]}{\sum_{j=1}^{M+1}\prod_{h=1}^{k_i}[\exp\left(\sum_{w=1}^{W}\phi_{hw}Z_{hw}(\boldsymbol{x_{ij}})\right)\prod_{s=1}^{S}(1 + \alpha_{hs} \exp\left(-(x_{ijs}/\beta_{hs})^2\right))]}\Big). \tag{2.12}$$

Under the homogeneity assumption in (2.10), maximizing (2.11) or (2.12) would be reduced to the constrained optimization problem with restriction $(\alpha_1 = ... = \alpha_K, \beta_1 = ... = \beta_K)$ or $(\alpha_{1s} = ... = \alpha_{Ks}, \beta_{1s} = ... = \beta_{Ks}, \forall s)$ respectively. The MLEs of PCMs can be obtained similarly. Standard errors of the parameter estimates can be calculated from the square root of the diagonal elements of the inverse of the Hessian matrix of the corresponding conditional likelihood, and then the 95% Wald-type confidence intervals (CI) can be constructed.

Parameter estimates and CIs can also be obtained using the profile likelihood. This approach reduces the number of independent parameters by expressing some of them as functions of the others, instead of dealing with all the parameters simultaneously. It is helpful in the presence of many parameters, such as in (2.11) and (2.12).

Remark 2.1: *Identifiability and Monte Carlo tests*. The likelihood based inference described above assumes that usual regularity conditions hold (Breslow et al. 1980). Under these regularity conditions, approximate CIs for the MLEs can be derived from the asymptotic multivariate normality of the MLEs and the estimated Hessian matrix. The likelihood ratio statistics for testing $H_0 : f(x) = 1$ has an asymptotic chi-squared distribution under the same regularity conditions. Diggle et al. (2000) pointed out that with an insufficient sample size, the log-likelihood surface of (α, β) may be far from quadratic and standard likelihood-based asymptotics are unreliable. Moreover, these models have an irregularity at the null hypothesis of $H_0 : f(x) = 1$, since $f(x) = 1$ corresponds to one of the two parameters of (α, β) equal to 0 with the other indeterminate, in the situation where there is no covariate adjustment. Monte Carlo tests can be used as an alternative. 1000 data sets can be simulated under the null and the observed values of the likelihood ratio statistics $LR = 2 \times (l(\hat{\alpha}, \hat{\beta}) - l(\alpha = 0 \text{ or } \beta = 0)) = 2(l(\hat{\alpha}, \hat{\beta}) - N \log(\frac{1}{M+1}))$ can be ranked among the 1000 simulated LR values. If the observed LR ranks *k*-th largest among 1000 simulated values, the p-value of the Monte Carlo test is k/1001 and the test is exact (Diggle et al. 2000).

Iteratively reweighted least square regression

Another alternative approach is IRLS regression. As the strata are mutually independent under the matched case-control design, it is not necessary to further consider the correlation between the residuals from different strata. Typically, one can write the non-linear regression model with binary response Y_i as

$$Y_i = p_i(\boldsymbol{x_i}, \boldsymbol{\theta}) + \varepsilon_i,$$

where Y_i is the observed binary response, $p_i(\boldsymbol{x_i}, \boldsymbol{\theta})$ is the predicted probability from the model for subject *i*, and $\varepsilon_i \sim N(0, \sigma^2)$ are independent and identically distributed random

errors, $i = 1, ..., N$. Under the conditional framework given there being a matched case-control pair at distance $\mathbf{x_i}$, we can treat each stratum as a single 'subject' with response $\sum_{j=1}^{M+1} I(Y_{ij} = k_i) = I(Y_{i1} = k_i)$ (assumed the first subject to be the case) and predicted probability $L_i^{k_i}$. One can further assume that the variance structure of the errors to be $\varepsilon_i \sim N(0, \sigma_k^2)$ for $\{i : k_i = k\}$, i.e., for all the strata where case response equals to k. Then the IRLS estimation can be realized by iteratively minimizing the weighted *SSE*

$$SSE(\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \sum_{k=1}^{K} \big[\sum_{i:k_i=k} \big(I(Y_{i1} = k) - L_i^k(\boldsymbol{x_i}, \boldsymbol{\theta})\big)\Sigma_k^{-1}\big(I(Y_{i1} = k) - L_i^k(\boldsymbol{x_i}, \boldsymbol{\theta})\big)\big], \tag{2.13}$$

where Σ_k is the pooled variance of errors from all strata where the case response equals k. In the initial step of IRLS, $\boldsymbol{\theta}$ is estimated by minimizing the weighted *SSE* with all $\Sigma_k^{(0)}$ set to identity. An estimate for $\Sigma_k^{(1)}$ is then calculated by $(1/df_k) \sum_{i:k_i=k} r_i^{(0)2}$, where the residuals $r_i^{(0)} = I(Y_{i1} = k_i) - L_{i1}^{k_i}(\boldsymbol{x_i}, \hat{\boldsymbol{\theta}}^{(\mathbf{0})})$ and df_k is the degree of freedom (the size of the set $\{i : k_i = k\}$ minus the number of parameters in the model). The estimated $\hat{\Sigma}_k^{(1)}$ are used as the weights in the next step of IRLS to minimize the weighted *SSE*. Parameter estimation is simply realized by iterating this process further, calculating updated estimates for Σ_k's, estimating the model parameters $\boldsymbol{\theta}$ with updated weights and iterating until convergence. The standard errors can be calculated from the Hessian matrix of the corresponding log likelihood

$$\sum_{k=1}^{K} \sum_{i:k_i=k} \big[-\frac{1}{2}\log(2\pi\sigma_k^2) - \frac{1}{2}\big(\frac{I(Y_{i1} = k) - L_i^k(\boldsymbol{x_i}, \boldsymbol{\theta})}{\sigma_k}\big)^2\big].$$

IRLS estimate and MLE were shown to be consistent and asymptotically normal under the assumption that the errors are normally distributed as $\varepsilon_i \sim N(0, \sigma_k^2)$ for $\{i : k_i = k\}$ (Gallant 2009).

For the three methods described above, instead of working directly on $(\alpha_{ks}, \beta_{ks})$ with a range of $(-1, \infty) \times (0, \infty)$, we performed unrestricted optimizations on the one-to-one

transformed parameters $(u_{ks}, v_{ks}) = (\log(1+\alpha_{ks}), \log(\beta_{ks}))$ that span the whole real plane, and then transformed the results back in terms of the original parameters $(\alpha_{ks}, \beta_{ks})$.

Bayesian approach

The Bayesian approach provides an alternative to the frequentist inferential strategies. A proper Bayesian approach would be to use the full likelihood and specify a prior distribution on the nuisance parameters $\boldsymbol{\rho} = (\rho_1, ..., \rho_N)$. However, the full likelihood approach would encounter the difficulty of prior specification and estimation of $\boldsymbol{\rho}$. One can use a marginal likelihood instead, that integrates out the nuisance parameters with respect to a random distribution. The equivalence between the use of conditional and marginal likelihoods for matched case-control study was discussed by Rice (2008). Diggle et al. (2000) pointed out that the conditional likelihood approach is consistent with the full likelihood approach for the binary outcome model with independent priors for $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$ (Diggle et al. 2000). Therefore, we proceed with the conditional likelihood as the basis for Bayesian inference.

The following sets of mutually independent prior distributions on $(\boldsymbol{u}, \boldsymbol{v}) = (u_{11}, ..., u_{KS}, v_{11}, ..., v_{KS})$ were primarily considered,

$$\begin{aligned} \log(1+\alpha_{ks}) = u_{ks} &\sim N(\mu_{u_{ks}}, \sigma^2_{u_{ks}}), \\ \log(\beta_{ks}) = v_{ks} &\sim N(\mu_{v_{ks}}, \sigma^2_{v_{ks}}), \quad k = 1, ...K; \quad s = 1, ...S, \end{aligned}$$

where the mean and variance of α_{ks} are $\mu_{\alpha_{ks}} = \exp(\mu_{u_{ks}} + \frac{1}{2}\sigma^2_{u_{ks}}) - 1$ and $\sigma^2_{\alpha_{ks}} = (\exp(\sigma^2_{u_{ks}}) - 1)\exp(2\mu_{u_{ks}} + \sigma^2_{u_{ks}})$ respectively. Similarly, $\mu_{\beta_{ks}} = \exp(\mu_{v_{ks}} + \frac{1}{2}\sigma^2_{v_{ks}})$ and $\sigma^2_{\beta_{ks}} = (\exp(\sigma^2_{v_{ks}}) - 1)\exp(2\mu_{v_{ks}} + \sigma^2_{v_{ks}})$. Both informative and noninformative (or vague) prior distributions were considered. For informative priors, based on our knowledge of roadway effects on asthma and the literature reviewed in introduction, the prior

distribution of α_{ks} was set with mean $\mu_{\alpha_{ks}} = 0.5$ and variance $\sigma^2_{\alpha_{ks}} = 0.25$ (thus $P(0.1 < \alpha_{ks} < 1.0) \approx 0.95$). For other types of health outcomes or pollution sources, different informative priors could be used. Given the fact that the point source effects on health outcomes (e.g. roadway effects on asthma) last only for a few hundred meters in most of the literature, prior distributions of β_{ks} were set with means $\mu_{\beta_{ks}} = 400$ and variance $\sigma^2_{\beta_{ks}} = 150$ (thus $P(50 < \beta_{ks} < 750) \approx 0.95$). For noninformative priors, the same mean $(\mu_{\alpha_{ks}}, \mu_{\beta_{ks}}) = (0.5, 400)$ with large variance $(\sigma^2_{\alpha_{ks}}, \sigma^2_{\beta_{ks}}) = (0.5, 400)$ were used for (α_{ks}, β_{ks}). It follows that $P(-0.2 < \alpha_{ks} < 2.0) \approx 0.95$ and $P(50 < \beta_{ks} < 1500) \approx 0.95$, which should contain the prior knowledge about (α, β). For the rest of this chapter, we focus on $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and primarily proceed using models without covariate adjustment.

A sensitivity analysis is performed by comparing the posterior distributions derived from various normal priors with the same means of $(\mu_{u_{ks}}, \mu_{v_{ks}})$ but different choices of $(\sigma^2_{u_{ks}}, \sigma^2_{v_{ks}})$. Wakefield and Morris (2001) suggested using independent Uniform prior distribution on (α, β) on the range of $(-1, \alpha_{max}) \times (0, \beta_{max})$ for the one point source binary model (2.1), where α_{max} and β_{max} are the maximum plausible values based on current epidemiological knowledge. This Uniform prior distribution on $(\alpha_{ks}, \beta_{ks})$ with different choices of α_{max} and β_{max} is also considered as part of the sensitivity analysis.

Since the full conditional distributions of the parameters do not follow a standard distributional form, the MCMC method is used to generate random draws from the posterior distributions. For two-parameter models such as the one point source homogeneous ACM, the random walk Metropolis-Hastings algorithm is used to generate a Markov chain which has the limit distribution equal to the target posterior distribution. For four(or more)-parameter models such as one point source ACM, computationally it is hard to draw simultaneously from the joint distribution using Metropolis-Hastings algorithm. Instead, we use a component-wise Metropolis-Hastings within Gibbs algorithm. The computa-

tional strategy corresponding to these MCMC algorithms is discussed in Appendix B.1. The convergence of these Markov chains are examined using Gelman and Rubin's convergence diagnostic (Gelman and Rubin 1992). In this study, the random walk Metropolis-Hastings or Metropolis-Hastings within Gibbs algorithm for the proposed models converge to their limit distributions after 2000-4000 runs. The chains have auto-correlations up to 20. Therefore, the chains are refined by choosing a common burn-in period of 5000, and a common thinning frequency of 20. These MCMC algorithms were performed for a length of $T = 45000$. After burn-in and thinning, the resulting Markov chains of length 2000 are treated as random draws from the target posterior distribution.

As a Bayesian counterpart to the Monte Carlo test, Bayes factors (Kass and Raftery 1995) are considered to test the null hypothesis that $H_0 : f(x) = 1$. The Bayes factor for comparing the current model M_1 to the null model M_0 is defined as the ratio of the posterior probability to the prior probability, which is given by

$$B = \frac{P(M_1|Y)/P(M_0|Y)}{P(M_1)/P(M_0)} = \frac{\int_\theta \pi(Y|\boldsymbol{\theta}, M_1)\pi(\boldsymbol{\theta}|M_1)d\boldsymbol{\theta}}{\int_\phi \pi(Y|\boldsymbol{\phi}, M_0)\pi(\boldsymbol{\phi}|M_0)d\boldsymbol{\phi}} = \frac{P(Y|M_1)}{P(Y|M_0)}.$$

The calculation of the Bayes factor B is not straightforward using MCMC. We used the importance sampling estimator $\frac{1}{T}\sum_{t=1}^{T}[l(\boldsymbol{\theta^t})\pi(\boldsymbol{\theta^t})/g(\boldsymbol{\theta^t})]$ as suggested by Diggle et al. (2000), where the prior distribution on $\boldsymbol{\theta}$ is used as the importance distribution $g(\boldsymbol{\theta})$, and $\boldsymbol{\theta^t}$ are sampled from $g(\boldsymbol{\theta})$. Kass and Raftery (1995) suggested calculating $2\log(B)$ as a Bayesian analogue of a log-likelihood ratio statistics or deviance. Values greater than 2 indicate increasing evidence against M_0: between 2 and 6 is 'positive' evidence, 6 to 10 is 'strong' and over 10 is 'very strong' evidence against M_0 (Diggle et al. (2000)Kass and Raftery (1995)). A number of alternatives can be found in DiCiccio et al. (1997).

2.3 A Simulation Study

Simulation scenarios

Two case subgroups ($K = 2$) and one control group, and up to two point sources are considered in the following simulation study. Specifically, four different settings of simulations are conducted where the true models are: 1) one point source PCM; 2) one point source ACM; 3) one point source homogeneous ACM; and 4) two point sources homogeneous ACM.

A large cohort of $L = 1,000,000$ people is generated initially. Two independent risk factors, age and gender, are included for this cohort, of which the distributions are set similar to those for the pediatric population of the Detroit Medicaid data source. Specifically, gender is generated from a Bernoulli distribution with probability 0.55 for being a male; age is generated from a piecewise Uniform distribution with a range of 2-18, and then rounded to integer values. The exposure variable, distance to the point source, is generated from a mixture distribution of Uniform and Gamma. Specifically, distances (in meters) from the first and second sources are generated from $0.15 \cdot Uniform(0, 500) + 0.85 \cdot Gamma(shape = 3, rate = 0.003)$ and $0.2 \cdot Uniform(0, 500) + 0.8 \cdot Gamma(shape = 3, rate = 0.005)$ respectively. Simulation studies are based on this fixed cohort with mutually independent covariates of age, gender and distances with distributions described above.

The disease status for the cohort would be different for different choices of distance-odds model or true parameter settings. For example, one point source ACM, the disease states ($k = 0, 1, 2$) are generated using the subject specific risk functions $\boldsymbol{p}(\boldsymbol{x})$ in (2.14) with certain fixed values of $(\alpha_1, \beta_1, \alpha_2, \beta_2)$. Specifically, the outcome for the *l*-th patient

Y_l is generated from the multinomial distribution with probabilities

$$
\begin{aligned}
P(Y_l = 0|x_l) &= \frac{1}{1 + \rho_{l1} f_1(x_l) + \rho_{l1}\rho_{l2} f_1(x_l) f_2(x_l)}, \\
P(Y_l = 1|x_l) &= \frac{\rho_{l1} f_1(x_l)}{1 + \rho_{l1} f_1(x_l) + \rho_{l1}\rho_{l2} f_1(x_l) f_2(x_l)}, \\
P(Y_l = 2|x_l) &= \frac{\rho_{l1}\rho_{l2} f_1(x_l) f_2(x_l)}{1 + \rho_{l1} f_1(x_l) + \rho_{l1}\rho_{l2} f_1(x_l) f_2(x_l)}, \quad l = 1, ..., L.
\end{aligned} \tag{2.14}
$$

The subject specific nuisance parameter for the *l*-th patient can be generated using $\rho_{lk} = \exp(b_{0k} + b_1 \times age_l + b_2 \times gender_l)$, $k = 1, 2$. The parameters $(b_{01}, b_{02}, b_1, b_2)$ can be obtained from the Detroit Medicaid data. Here $b_1 = -0.05$ and $b_2 = 0.3$ are used. The intercepts b_{01} and b_{02} can be varied within a range of $(-2.0, -0.5)$ to generate different desired disease prevalence. Typically, about 20% subjects of the cohort are generated as cases, of which all disease sub-categories have roughly the same proportion ($k = 1, \approx 10\%$; $k = 2, \approx 10\%$). After the disease status is generated for the cohort, $R = 500$ matched case-control data sets are then generated, each with N 1:1 matched pairs. Different sample sizes $N = 500$, 1000 and 2000 are also considered. Specifically, for each of the R matched case-control data sets, N cases are randomly drawn from the cohort, and then they are randomly matched with controls by age (within 2 years) and gender. Covariate adjustment was not considered in the simulation study since both covariates of age and gender are matched.

Under each model setting, parameter estimates with 95% CIs are calculated using MLE, profile likelihood and IRLS. Due to the identifiability problem of the likelihoods for the proposed models, there are a few runs ($< 5\%$) that fail to converge, or converge for the point estimates but can not obtain CIs (for example, failure to invert the Hessian matrix using maximum likelihood method). The non-converged data sets among the $R = 500$ ones were removed. The simulation results are summarized on the remaining R' data sets where all three frequentist methods converge. The R' estimates are summarized in terms

of relative bias (e.g. relative bias for a parameter θ is $(\frac{1}{R'}\sum_{i=1}^{R'}\hat{\theta}_{(i)} - \theta_{true})/\theta_{true} \times 100\%$), MSE (e.g. $MSE = \frac{1}{R'}\sum_{i=1}^{R'}(\hat{\theta}_{(i)} - \theta_{true})^2$) and coverage probability (the proportion that the 95% CIs cover the true value is calculated as an *ad hoc* estimate of the true coverage probability among these R' runs). For the Bayesian approach, the posterior mode as well as 95% highest posterior density (HPD) interval are estimated based on 2000 draws (after burn-in and thinning) from the posterior distribution. Since the posterior distributions of α and β are both positively skewed (a heavy right tail for β), the posterior mean is not used. In order to compare with the frequentist results such as MLE, the posterior mode is used instead of the median because the posterior mode asymptotically converges to MLE. The R' posterior modes are summarized in terms of relative bias and MSE for the same R' data sets. The coverage probability is calculated as the proportion of times that the 95% HPD intervals cover the true value.

Simulation results

A summary of the simulation results comparing convergence rate, relative bias and coverage probability by different methods and by different sample sizes is shown in Table 2.1, for the four distance-odds models (i.e. one point source PCM, ACM and homogeneous ACM, and two point sources homogeneous ACM). The MSE comparison is summarized in Figure 2.1. Since the three frequentist methods of MLE, Profile likelihood and IRLS regression provide very similar and consistent results, we primarily focus on the difference between the broad class of frequentist and Bayesian approaches which is described below in terms of convergence, relative bias, MSE and coverage probability separately. Additionally, the following results hold for $\boldsymbol{\alpha}$'s and $\boldsymbol{\beta}$'s. The complete numerical simulation results can be found in Appendix A.1.

Convergence: For all four distance-odds models with a large sample size such as $N =$

2000, the frequentist methods perform well in terms of convergence with a joint convergence rate $R'/R > 90\%$. Typically, less than 5% of runs failed to converge for each of the three frequentist methods. With a decreased sample size of $N = 500$, the 90% joint convergence rate remains for the two homogeneous models. However, failures increase to 30% for one point source PCM and ACM using frequentist methods. Thus, the simulations for these two models were performed and presented for a sample size of $N = 1000$ in Table 2.1, where a joint convergence rate of 85% occur using frequentist methods. In the Bayesian approach we numerically assessed the convergence of the posterior chains by the Gelman-Rubin convergence diagnostic (Gelman and Rubin 1992). No problems were detected either numerically or via examining the trace plots in our limited simulation study. The MCMC method does not require the usual regularity conditions (Breslow et al. 1980) or any asymptotic normality assumption, and it yields exact posterior distributions for all sample sizes. It also avoids the identifiability issue, but needs a careful choice of the covariance matrix of the proposal distribution because of the strong correlations among the model parameters.

Relative bias: When $N = 2000$, low relative biases (with range $(-9.2, 10.7)\%$ for $\boldsymbol{\alpha}$'s and $(-2.9, 4.2)\%$ for $\boldsymbol{\beta}$'s) are observed for both frequentist and Bayesian methods for all models with different choices of true parameter settings (shown in Table 2.1, numerical details can be found in Appendix A.1). Thus, both methods have performed well with large sample size in terms of relative bias. For smaller sample sizes ($N = 500$ for the two homogeneous models; $N = 1000$ for one point source PCM and ACM), relative biases of $\boldsymbol{\alpha}$ are usually as high as 25%, while relative biases of $\boldsymbol{\beta}$ are still well controlled ($< 5\%$, except few extreme setting). Note that, estimates of $\boldsymbol{\alpha}$ are biased upwards (Table 2.1) using frequentist methods with these small sample sizes, while Bayesian methods do not suffer as much. The above results are consistent across inference methods for each model

as shown in Table 2.1.

Mean squared error: When the sample size $N = 2000$, the MSEs are consistent across methods for each distance-odds models with different true parameters. Figure 2.1 shows the MSEs corresponding to each method with smaller sample sizes of $N = 500$ or 1000. The three frequentist approaches using MLE, profile and IRLS method show very similar MSE values, while the Bayesian approach shows consistently lower MSEs than frequentist approach for each distance-odds model regardless of true parameters values. Note that, for the Bayesian approach, the MSEs derived from informative priors are much lower than those from noninformative (vague) priors for each setting as expected. Thus, if prior knowledge is available, it should be used to enhance precision for these distance-odds models.

Coverage probability: In Table 2.1, when $N = 2000$, the coverage probabilities are around 95% for all the models and methods in our simulation study. For smaller sample sizes of $N = 500$ or 1000, the coverage probabilities fall below the nominal level for some parameter settings, however, they are still around 95% on average (shown in Table 2.1, numerical details can be found in Appendix A.1). Note that these percentages are estimated based on the R' data sets where all three frequentist methods converge. In addition, the Bayesian approach provides comparable percentages based on all $R = 500$ data sets. Therefore, it is more stable than the frequentist methods in terms of coverage probability and convergence.

In summary, Bayesian methods, especially incorporated with prior knowledge, have advantages in terms of estimation stability and precision for the proposed non-linear distance-odds models with multiple disease sub-types.

2.4 Application: Data analysis for the DAMAT Study

The Detroit asthma morbidity, air quality and traffic study describes a population-based matched case-control analysis investigating associations between acute asthma outcomes and proximity of residence to major roads in Detroit, MI. We examined the pediatric population (2-18 years of age) served by Medicaid for the study period 2004 through 2006. The Medicaid data provide the most complete and readily available source of healthcare utilization across Detroit. The population consists mainly of African American children from lower income families, and is considered a high risk population for asthma-related events. The data included an encrypted Medicaid identifier, age, sex, race/ethnicity, utilization dates and diagnostic codes for inpatient admissions and emergency department visits, and geo-coded home residence at the time of each health care visit. To ensure a full claims history, the study population was restricted to those with continuous Medicaid enrollment (more than 11 months in each year), full Medicaid coverage, and no other insurance. Asthma cases were identified as all children who made at least one asthma claim during the three-year study period, indicated by primary diagnostic code 493.X (International Classification of Diseases, 9th Revision, Clinical Modification). Controls were defined as children whose primary diagnosis was injury or poisoning. Each asthma case was matched with one control based on gender, race and age (within 2 years). Asthma cases were further grouped into multiple disease categories ($K = 2$), based on the frequency of acute asthma outcomes ($Y = 2$: claimants with 2 or more asthma claims; $Y = 1$ claimants with exactly 1 asthma claim). Details on the descriptive analysis of this data set can be found in Li et al. (2011).

The geo-coded residence information was used to estimate the distance to major roads in Detroit, defined as state and interstate freeways and major arterials with annual average

daily traffic (AADT) flows exceeding 50,000 and 20,000 vehicles per day, respectively. The freeways and the arterials are considered as the first and second point sources respectively. Shape files providing coordinates of road centerlines were obtained from the Southeast Michigan Council of Governments. These files and the geo-coded claim data were merged into ARCGIS 9.3 to determine the proximity to each major road. Due to confidentiality concerns, claim locations were reported only to the closest 10 m. The road centerline does not account for the width of the highway and median strip, if any, which can exceed 30 m for sections of some freeways. Taken together, these factors suggested that differences on the order of at least 20 to 50 m would be meaningful.

Separate analyses were performed for one and two point source(s) models. For one point source (freeways) models, the study region was restricted to 1,000 m buffer of freeways, which consisted of 2669 1:1 matched case-control pairs. For two point sources (freeways and arterials) models, the study region was restricted to 1,000 m buffer of freeways or arterials, which consisted of 4081 1:1 matched case-control pairs. Figure 2.2 illustrates the natural spline fit and 95% confidence band for the relationship between distance to roadways and odds of being an asthma claimant, using a CLR model with only spline of distance as its argument. These plots provide an exploratory analysis of the data which indicate increasing risk with proximity to both types of roads, where the freeways appear to have stronger effects. There may be a threshold distance beyond which the roadway effect vanishes. The increase of odds at $600m$ of freeways is not statistically significant, which could be an artifact of the smoothing parameter ($df = 3$ in the natural spline).

Method comparison: The frequentist methods of MLE, profile likelihood and IRLS provide similar point estimates and CIs with essentially the same Akaike information criterion (AIC) values for each distance-odds model. Thus, we primarily discuss results as frequentist method (MLE as demonstration) versus Bayesian method in the main text. Table 2.2

shows the parameter estimates and 95% CIs using likelihood method and posterior modes with 95% HPD intervals using Bayesian methods, for one point source models. Note that these log-likelihood surfaces are not far from quadratic in shape given the large sample size of 2669 asthma cases in the DAMAT study. Note also that the contour lines near $u = 0$ (or equivalently $\alpha = 0$) are almost vertical, which implies the identifiability issue that a wide range of β can provide the same value of likelihood values. Fortunately, the peaks of the likelihood surfaces are not close to the null for these one point source models. For the Bayesian method, the locations of the posterior modes for one point source models are close to each other for the two prior choices for each parameter under each model. Posterior densities of $\boldsymbol{\beta}$ are highly right-skewed, especially for noninformative prior distribution with much wider HPDs than those derived from informative priors (shown in Table 2.2). Thus, the frequentist likelihood based inference method or a noninformative Bayesian method should be avoided for these distance-odds models in presence of well elicited prior knowledge.

Model selection: Generally, the distance-odds models are selected *a priori* in the study design stage. For example, different choices of the numbers of point sources would provide different study regions with different sample sizes. The choice between PCMs and ACMs can also be considered *a priori* based on the interest of nominal or ordered disease sub-classifications. Model selection can also be based on AICs for frequentist method or DICs for Bayesian method. For example, ACM (homogeneous) has the smallest AIC value among the four one point source models as shown in Table 2.2. However, the differences among these AICs are very small and of little practical concern. In this case, all these one point source models fit almost equally well. For both informative and noninformative priors, one point source PCM and ACM have similar and relatively lower DIC values than the other two models. There is evidence that the more sophisticated models that allow

different functional forms of odds between case sub-types are preferred even after penalizing for the additional number of parameters using the Bayesian approach. Therefore, a PCM (smallest DIC) with informative priors is the preferred approach among all one point source models for the DAMAT study (different numbers of point sources with different sample sizes are not directly comparable). Similarly, Table 2.3 shows the corresponding results for the two point sources binary model and homogeneous ACM, where the later one with an informative prior Bayesian approach is preferred.

Estimation and interpretation: Table 2.2 shows the parameter estimates and 95% CIs using MLE, and posterior modes with 95% HPD intervals using Bayesian methods, for the one point source models (binary/ACM/PCM). Generally, the point estimates of $\hat{\alpha}$ and $\hat{\beta}$ lay within $0.1 - 0.4$ and $100 - 300$ respectively for the one point source models, which implies that the roadway effect on asthma only lasts up to a few hundred meters and that the increase in risk is modest. Take the one point source PCM that has the smallest DIC as an example, the MLE (or Posterior Mode) $\hat{\alpha}_2 = 0.39(0.32)$ is slightly larger than $\hat{\alpha}_1 = 0.21(0.25)$ as shown in Table 2.2. It implies that, at the point source, the odds of asthma for claimants with 2 or more claims ($k = 2$) versus controls is slightly higher than the odds for claimant with exactly 1 claim ($k = 1$) versus controls. Table 2.3 shows the results for two point sources models. In general, we have $\hat{\alpha}_{11} > \hat{\alpha}_{12}$ and $\hat{\beta}_{11} > \hat{\beta}_{12}$, which implies the odds of asthma at freeways is higher than the odds at arterials, and the freeways effects last longer than arterials. Figure 2.3 shows the estimated distance-odds functions $\hat{\boldsymbol{f}}_{\boldsymbol{k}}$ for the one point source PCM, using MLE and Bayesian method with informative priors. Note that the Bayesian method with prior knowledge provides consistently higher estimates of $\boldsymbol{f}_{\boldsymbol{k}}$ than MLE. For both case subgroups, $\hat{\boldsymbol{f}}_{\boldsymbol{k}}$ deceases rapidly within $0 - 300$ meters, and then the roadway effect on asthma lasts up to 400 meters off freeways using MLE method and 600 meters using Bayesian method respectively. The 95% credible

regions are above unity up to a distance of 350 meter. Note that the MLE of $\boldsymbol{f_k(\alpha, \beta)}$ is estimated by plugging in the MLE of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ using their invariant property; the posterior distribution of $\boldsymbol{f_k(\alpha, \beta)}$ is estimated by draws from the posterior distribution of $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ for fixed grid values of distance x (every half meter). Note also that, for interval estimates of a function of parameters, the 95% Bayesian credible region can be directly obtained from the draws, however, the calculation of the frequentist confidence bands for the MLE of $\boldsymbol{f_k(\alpha, \beta)}$ is not straight forward, this requires the Delta theorem (calculation of the first and second derivatives of the complex likelihood function) and relies on asymptotic properties needing a large sample size.

Table 2.4 shows the p-values of the Monte Carlo test and Bayes factors for testing $H_0 : f(x) = 1$ for one and two point source(s) distance-odds models. Evidence of associations ($H_1 : f(x) > 1$) is found for most models using the MC test (P-value< 0.05) or Bayes factors ($B > 2$). Strongest associations are found for PCM among one point source models and for homogeneous ACM among two point sources models respectively, which is consistent with the results in Tables 2.2 and 2.3.

Sensitivity analysis: The results in Tables 2.2, 2.3 and 2.4 show consistency for different choices of the distance-odds models under a matched case-control study. Similar conclusion can be drawn using these models that there is evidence of the roadway effect on asthma, and that the effect is modest and only lasts up to a few hundred meters. As a sensitivity analysis of the prior specification, posterior densities are derived and compared from different choices of prior distributions for the one point source PCM. For normal priors on (u, v) with different variances (σ_u^2, σ_v^2), the posterior modes are close to each other for each parameter under each model (shown in Appendix A.1). However, the posterior modes are sensitive to the choice of α_{max} and β_{max} using Uniform priors on $(-1, \alpha_{max})$ and $(0, \beta_{max})$. When α_{max} and β_{max} are large, these Uniform priors still put equals weights

on the whole range of $(-1, \alpha_{max})$ and $(0, \beta_{max})$ that may overly weight the upper extreme values. Wakefield and Morris (2001) have also pointed out the influence of the Uniform priors, which reflects the fact that there is little information in the likelihood due to sparsity of data in the upper extremes. Thus the parameterization (u, v) with normal priors appear to be more robust.

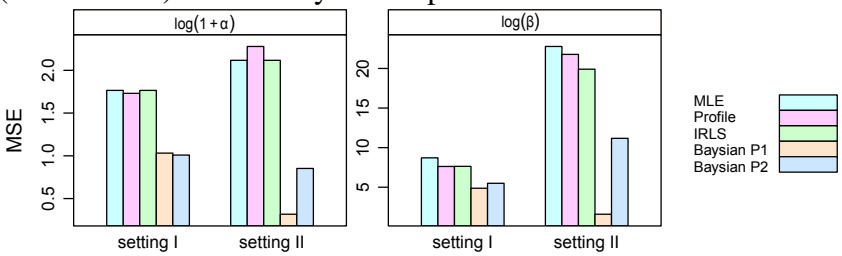
2.5 Discussion

In this chapter, we extended the distance-odds model of Diggle et al. (2000) to models where there are sub-types within cases under a matched case-control design. The extension to sub-classification within cases is non-trivial with these non-linear odds functions under a matched design. Maximum likelihood, profile likelihood, IRLS and a Bayesian approach using MCMC methods were evaluated under the proposed models. We compared these methods via an extensive simulation study evaluating frequentist properties such as relative bias, MSE and coverage probability, and showed that Bayesian methods have advantages in terms of estimation stability, precision and interpretation. The Bayesian methods are able to yield direct HPD for complex non-linear distance-odds functions, and does not require large sample approximation. There is no simulation study in literature that compares the convergence, relative bias, MSE or coverage probability for these point source models, even for the basic binary outcome model. The proposed models and methods are applied to a population-based matched case-control study investigating associations between acute asthma outcomes and proximity of residence to major roads by analyzing Medicaid claims data for the pediatric asthma population in Detroit, MI, from 2004-2006. We also perform a sensitivity analysis to investigate how the choice of distance-odds models and specification of the prior distributions affect the results. Typically, the results were consistent for different choices of models and normal prior distributions on the transformed parameters

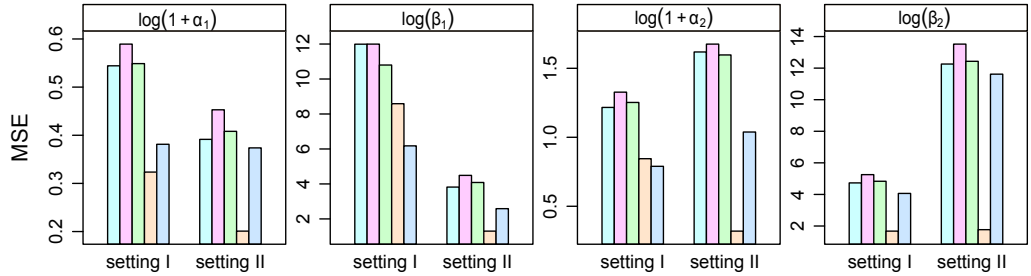
for the DAMAT study.

The current study has several limitations which may lead to future research. The extension of the non-linear distance-odds model to the proportional odds model setting is not considered in the study, which is most commonly used for ordered data. We realize that the conditional likelihood does not apply to this model due to the nuisance parameters remaining in the non-linear odds functions. The prospective-retrospective conversion for case-control data is only valid for a multiplicative intercept model. Moreover, the way of modeling the point source as a function of distance assumed that the distance is the only factor that matters but not the other factors at the position where the individual lives (e.g. wind speed, wind direction, barrier not allowing the pollution to travel). Two individuals could live at the same distance from the freeway but in two different residence and experience different amount of air pollution simply because of the way the wind usually blows. One way to handle this problem is to replace the odds ratio function with a Gaussian correlation function that is not isotropic. Examples of anisotropic correlation functions can be found in Banerjee et al. (2003) and Baddeley et al. (2010). Many literatures (HEI 2010) have addressed the issues of modeling near-road concentrations that take wind speed and wind direction into account, e.g. a reduced-form dispersion model (Batterman et al. 2010). However, in this project we only considered point source modeling as a function of distance only, which can be assumed as a model where the average effect of all the other factors (e.g. the wind direction across time for a three-year period) are canceled out. These issues remain to be explored in future research.

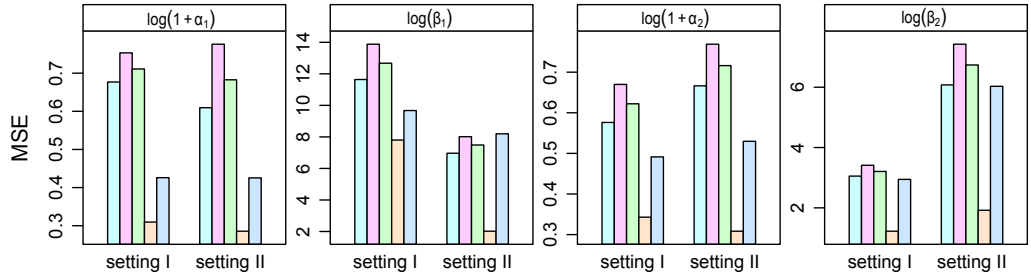
Figure 2.1: Mean squared errors for two settings of true parameter values under various distance-odds models, using MLE, profile likelihood, IRLS and Bayesian methods with $R = 500$ simulations. Bayesian P1 and P2 refer to two choices of prior distributions; Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$. Y-axis (MSE values) is scaled by a multiplier of 100.



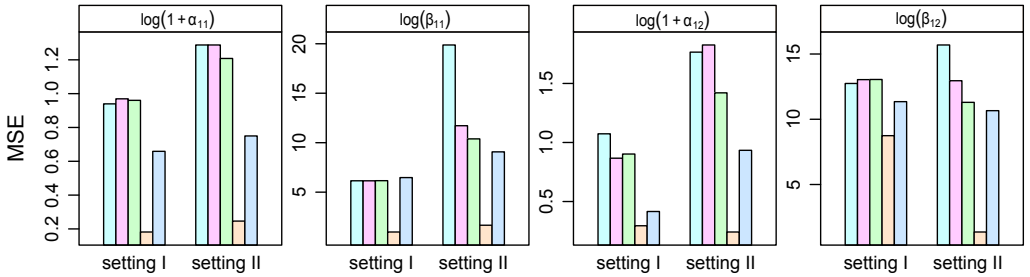
(a) **one point source homogeneous adjacent category model.** Sample size $N = 500$; Settings I and II refer to $(\alpha, \beta) = (0.7, 300)$ and $(0.4, 500)$ respectively.



(b) **one point source adjacent category model.** Sample size $N = 1000$; Settings I and II refer to $(\alpha_1, \beta_1, \alpha_2, \beta_2) = (0.3, 300, 0.7, 500)$ and $(0.4, 500, 0.4, 500)$ respectively.



(c) **one point source polychotomous category model.** Sample size $N = 1000$; Settings I and II refer to $(\alpha_1, \beta_1, \alpha_2, \beta_2) = (0.3, 300, 0.7, 500)$ and $(0.4, 500, 0.4, 500)$ respectively.



MSEs versus inference methods for two settings of true parameter values

(d) **two point sources homogeneous adjacent category model.** Sample size $N = 500$; Settings I and II refer to $(\alpha_{11}, \beta_{11}, \alpha_{12}, \beta_{12}) = (0.5, 500, 0.3, 300)$ and $(0.4, 500, 0.4, 500)$ respectively.

Figure 2.2: Estimated natural spline terms of distance showing the distance-odds relationships for asthma claimants versus controls, using (binary) conditional logistic regression model with spline of distance as its argument. The solid lines show the point estimates; the dashed lines show the 95% confidence bands.

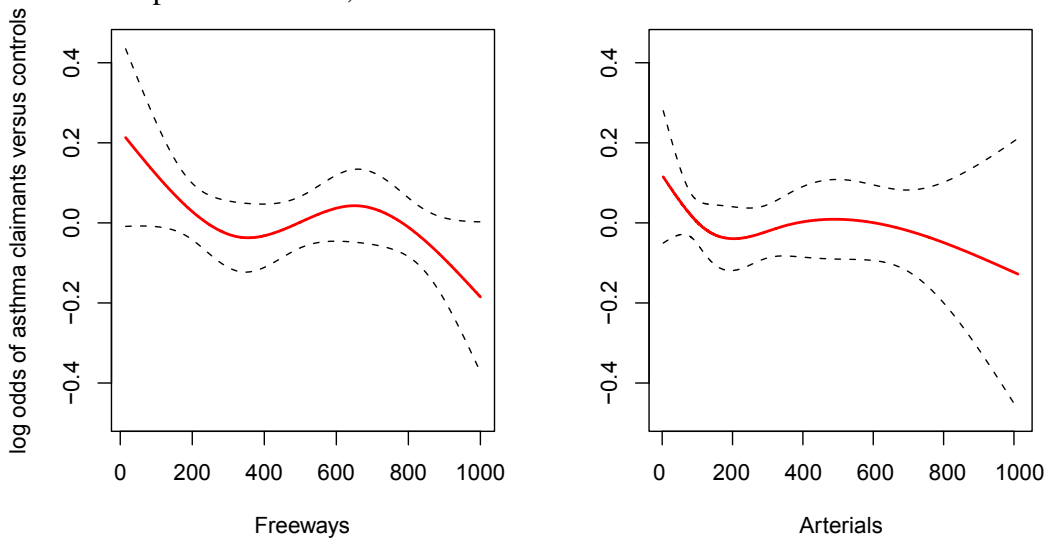


Figure 2.3: Estimated distance-odds functions for the one point source polychotomous category model. The solid blue line shows the MLE of the odds function; the solid red line shows the Bayesian posterior mode estimate with 95% credible region (dashed lines). Parameters of prior distribution used are $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$.

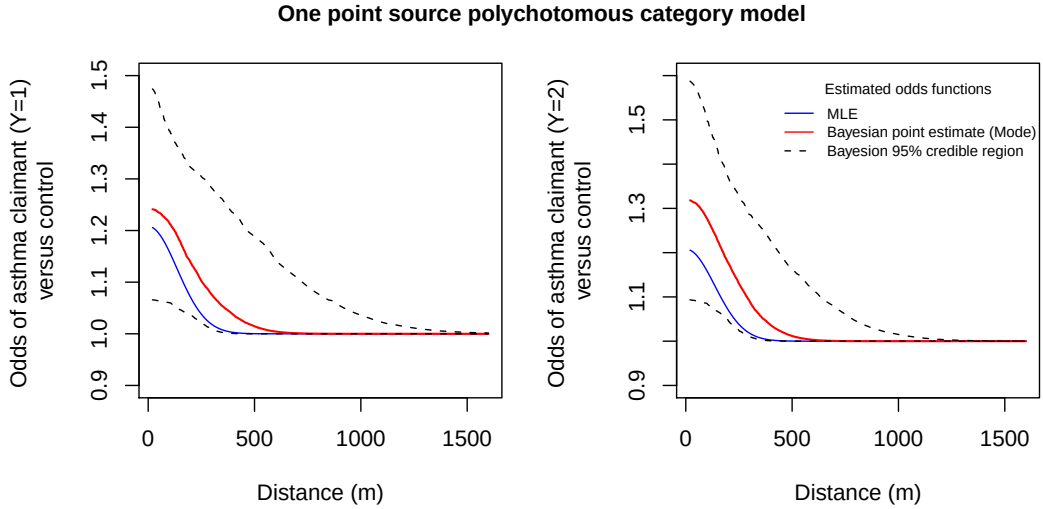


Table 2.1: Summary of the simulation results in terms of convergence rate, relative bias and coverage probability comparing frequentist and Bayesian methods using different sample sizes.

		Frequentist method			Bayesian method		
		Lack of	RB[a](%)	CP[a](%)	Lack of	RB (%)	CP (%)
N=2000		convergence[a]	α β		convergence	α β	
One point source	ACM (H)[b]	3%	$(-0.5, 6.9), (-0.4, 1.3)$	$93.8(91, 97)$	None noted	$(-8.3, 7.6), (-0.4, 3.9)$	$93.2(81, 98)$
	ACM	8%	$(-3.4, 9.6), (-1.9, 1.7)$	$95.2(90, 98)$	None noted	$(-9.2, 7.7), (-1.1, 3.2)$	$94.1(92, 98)$
	PCM	9%	$(-0.2, 4.6), (-0.9, 2.1)$	$93.7(91, 98)$	None noted	$(-4.2, 4.9), (-0.9, 4.2)$	$93.6(89, 97)$
Two point source	ACM (H)	5%	$(0.4, 10.7), (-0.3, 0.6)$	$94.6(89, 97)$	None noted	$(-9.2, 9.2), (-2.9, 2.8)$	$94.9(84, 99)$
N=500 or 1000[c]							
One point source	ACM (H)	8%	$(9.0, 25.6), (-0.8, 1.7)$	$93.2(84, 99)$	None noted	$(-15.8, 17.6), (-0.8, 6.0)$	$94.3(67, 99)$
	ACM	15%	$(3.3, 17.5), (-0.6, 1.4)$	$93.6(80, 99)$	None noted	$(-20.1, 15.7), (-4.4, 8.3)$	$95.5(86, 100)$
	PCM	13%	$(1.2, 14.9), (-0.2, 0.8)$	$93.6(89, 100$	None noted	$(-14.8, 14.1), (-3.4, 4.1)$	$95.5(88, 99)$
Two point source	ACM (H)	10%	$(10.0, 23.1), (-0.4, 0.8)$	$94.5(88, 99)$	None noted	$(-24.3, 18.9), (-4.6, 4.2)$	$95.6(81, 100)$

[a] Lack of convergence: mean of the none-convergence rates $(R - R')/R$ across parameter settings; RB: range of the **relative biases** across parameter settings; CP: mean and the range of the **coverage probabilities** across parameter settings.

[b] homogeneous adjacent-category model

[c] $N = 500$ for the two homogeneous models; $N = 1000$ for one point source PCM and ACM.

Table 2.2: Parameter estimates with 95% confidence intervals for **one point source models** using MLE, profile likelihood and IRLS methods; and posterior modes with 95% highest posterior density (HPD) credible intervals using MCMC.

MLE[a]	Binary Model	α	β			AIC
	Estimate	0.258	174.1			3699.9
	CI[a]	(-0.042, 0.558)	(55.7, 292.4)			
	ACM (Homogeneous)	α_1	β_1			
	Estimate	0.188	168.8			3699.8
	CI	(-0.023, 0.398)	(57.8, 279.8)			
	ACM (General)	α_1	β_1	α_2	β_2	
	Estimate	0.215	176.0	0.130	153.4	3703.7
	CI	(-0.126, 0.557)	(41.6, 310.4)	(-0.484, 0.744)	(87.5, 394.2)	
	PCM	α_1	β_1	α_2	β_2	
	Estimate	0.208	191.8	0.392	154.1	3703.6
	CI	(-0.118, 0.534)	(9.1, 374.6)	(-0.242, 1.025)	(26.5, 281.7)	
Bayesian P1[a]	Binary Model	α	β			DIC
	Posterior mode	0.247	228.6			3686.2
	Posterior median	0.289	290.7			
	CI (HPD)[a]	(0.034, 0.487)	(121.0, 592.1)			
	ACM (Homogeneous)	α_1	β_1			
	Posterior mode	0.177	182.7			3686.8
	Posterior median	0.156	202.5			
	CI (HPD)	(0.025, 0.361)	(118.1, 550.0)			
	ACM (General)	α_1	β_1	α_2	β_2	
	Posterior mode	0.194	192.5	0.242	222.5	3675.3
	Posterior median	0.244	287.2	0.261	326.8	
	CI (HPD)	(0.004, 0.461)	(125.5, 667.8)	(-0.072, 0.505)	(116.7, 623.3)	
	PCM	α_1	β_1	α_2	β_2	
	Posterior mode	0.246	231.3	0.320	259.0	3674.9
	Posterior median	0.298	256.8	0.366	269.7	
	CI (HPD)	(0.028, 0.514)	(113.3, 737.8)	(0.049, 0.649)	(115.3, 602.6)	
Bayesian P2[a]	Binary Model	α	β			DIC
	Posterior mode	0.285	152.7			3682.7
	Posterior median	0.203	398.2			
	CI (HPD)	(0.027, 1.308)	(154.6, 1401.2)			
	ACM (Homogeneous)	α_1	β_1			
	Posterior mode	0.192	160.5			3683.2
	Posterior median	0.216	395.2			
	CI (HPD)	(0.005, 1.086)	(79.3, 1263.2)			
	ACM (General)	α_1	β_1	α_2	β_2	
	Posterior mode	0.212	177.5	0.162	172.5	3670.4
	Posterior median	0.312	425.7	0.188	256.7	
	CI (HPD)	(-0.039, 0.896)	(96.5, 1347.8)	(-0.181, 0.840)	(39.5, 1123.9)	
	PCM	α_1	β_1	α_2	β_2	
	Posterior mode	0.258	243.2	0.286	147.0	3669.4
	Posterior median	0.346	566.4	0.367	218.9	
	CI (HPD)	(-0.056, 0.822)	(76.5, 1490.9)	(-0.080, 0.818)	(44.9, 1267.8)	

[a] MLE: maximum likelihood estimate; CI: confidence/credible interval; HPD: highest posterior density; Bayesian P1 and P2 refer to two settings of prior choice; Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$.

Table 2.3: Parameter estimates with 95% confidence intervals for **two point sources models** using MLE, profile likelihood and IRLS methods; and posterior modes with 95% highest posterior density (HPD) credible intervals using MCMC.

		First point source		Second point source		
MLE[a]	Binary Model	α_{11}	β_{11}	α_{12}	β_{12}	AIC
	Estimate	0.228	309.2	-0.098	180.5	5657.5
	CI	(-0.177, 0.663)	(97.3, 575.8)	(-0.420, 0.223)	(17.4, 376.2)	
	ACM (Homogeneous)	α_{11}	β_{11}	α_{12}	β_{12}	
	Estimate	0.179	283.6	-0.134	114.8	5656.0
	CI	(0.001, 0.360)	(68.4, 535.1)	(-0.357, 0.093)	(6.5, 233.2)	
Bayesian P1[a]	Binary Model	α_{11}	β_{11}	α_{12}	β_{12}	DIC
	Posterior mode	0.280	257.9	0.061	270.0	5609.1
	Posterior median	0.304	302.1	0.089	300.1	
	CI (HPD)[a]	(0.127, 0.462)	(171.9, 533.2)	(-0.072, 0.200)	(132.1, 543.7)	
	ACM (Homogeneous)	α_{11}	β_{11}	α_{12}	β_{12}	
	Posterior mode	0.205	294.0	0.019	261.4	5593.8
	Posterior median	0.212	360.2	0.021	340.4	
	CI (HPD)	(0.075, 0.354)	(155.0, 509.8)	(-0.083, 0.122)	(143.1, 633.2)	
Bayesian P2[a]	Binary Model	α_{11}	β_{11}	α_{12}	β_{12}	DIC
	Posterior mode	0.248	327.5	0.007	182.3	5604.6
	Posterior median	0.303	430.2	0.011	434.6	
	CI (HPD)	(0.069, 0.474)	(122.9, 627.2)	(-0.131, 0.134)	(75.1, 1225.7)	
	ACM (Homogeneous)	α_{11}	β_{11}	α_{12}	β_{12}	
	Posterior mode	0.186	228.0	-0.006	149.4	5595.2
	Posterior median	0.222	340.4	0.011	480.9	
	CI (HPD)	(0.051, 0.354)	(129.0, 645.8)	(-0.120, 0.108)	(70.1, 1243.2)	

[a] MLE: maximum likelihood estimate; CI: confidence/credible interval; HPD: highest posterior density; Bayesian P1 and P2 refer to two settings of prior choice; Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$.

Table 2.4: Monte Carlo test p-values and Bayes factors $2\log(B)$ for the null hypothesis that $H_0 : f(x) = 1$ for various point source(s) models. Bayesian P1 and P2 refer to two settings of prior choice; Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$.

Models	MC test p-values	Bayes factors P1	P2
One point source			
Binary model	0.04	3.52	2.89
ACM (Homogeneous)	0.06	4.32	3.41
ACM (General)	0.02	6.29	6.16
PCM	< 0.01	7.12	6.04
Two point source			
Binary model	0.04	3.11	2.57
ACM (Homogeneous)	< 0.01	6.69	5.98

CHAPTER III

Bayesian analysis of time-series data under case-crossover designs

3.1 Introduction

Case-crossover design, originally proposed by Maclure (1991), has been widely used to study the effect of short-term exposure on the risk of acute adverse health events, such as temperature on mortality (Basu et al. 2005) and ambient air pollution on asthma (Li et al. 2011). Under this design, exposure at the event time of each case is compared to exposure at some referent times (times within a certain period where the same case did not experience any event). For each case, a 'referent window' is defined as a set of time points consisting of the event time and all referent times for the same case. It can be viewed as a hybrid of case-control (comparing exposure distribution of cases and controls through a retrospective design) and crossover (the case serves as its own control) studies. The case-crossover design inherently controls for time-invariant confounders (e.g. gender and race) by making within-person comparisons, and controls for potential time-varying confounders (e.g. seasonal trends) by proper choice of the referent times. The time-stratified case-crossover design divides time *a priori* into disjoint strata, uses the event time to determine which stratum is selected, and selects all or a sub-sample of the remaining times in the stratum as referent times for a given case (Janes et al. 2005a). For example, time stratum based on the same day of the week in the same calender month that

controls for confounding due to day of the week, seasonal and long-term effects is often recommended (Janes et al. 2005a).

The design and analytic issues related to the referent time selection have been comprehensively discussed (Lumley and Levy (2000); Levy et al. (2001); Janes et al. (2005a); Janes et al. (2005b); Mittleman (2005)). The traditional approach for analyzing case-crossover data is to treat them as coming from a matched case-control structure, where each stratum consists of exposures at event and referent times of a given case. A conditional logistic regression (CLR) is routinely used to obtain estimates of the underlying risk ratio parameters. In terms of referent time selection, a 'non-localizable' design (Janes et al. 2005a) is a case-crossover design for which the CLR estimating equation under the choices of referent times is biased, such as unidirectional (Maclure 1991), bidirectional (Navidi 1998) and symmetric bidirectional designs (SBD) (Bateson and Schwartz 1999). The bias has been termed 'overlap bias' (Lumley and Levy 2000). In contrast, a 'localizable' design (Janes et al. 2005a) is a case-crossover design for which there exists an unbiased CLR estimating equation, such as the time-stratified design (TSD) (Janes et al. 2005a) and semi-symmetric bidirectional design (Navidi and Weinhandl 2002). Appendix A.2 Figure 1 shows several illustrations of common referent time selection strategies. The TSD is generally preferred compared to any of the alternatives thus far proposed (Janes et al. (2005b); Mittleman (2005)). Based on a 2010 review article (Carracedo-Martínez et al. 2010), though 42% of case-crossover studies during 1999-2008 used SBD, the TSD has become the most popular design since 2005.

An alternative analysis of such exposure and event series data is to use a standard time-series analysis. Lu and Zeger (2007) have shown that the traditional CLR approach to analyze case-crossover data can be viewed as a time-series analysis with an underlying log-linear model of a specific form. This equivalence has also been noted in special cases

by Levy et al. (2001) and by Janes et al. (2005a).

Bayesian data analysis under case-crossover designs appear to be non-existent in the literature though there is substantial work on Bayesian modeling of matched case-control data (Ghosh and Chen (2002); Sinha et al. (2004)). It is true that the use of CLR remains identical in the two contexts for certain 'localizable' designs. However, the assumptions and the data structure make the statistical points of discussion distinct in a case-crossover study compared to a matched case-control study under a Bayesian paradigm. In this chapter we consider a comprehensive treatment of the problem starting with some posterior equivalence results, followed by alternative Bayesian proposals beyond using CLR as the basis for inference in case-crossover studies.

This chapter is structured as follows. In section 3.2, we describe the disease-exposure association model, underlying assumptions and two potential likelihood formulations, the conditional and the full likelihood, under the case-crossover design. We then consider equivalence results analogous to Lu and Zeger (2007) in a Bayesian framework under both formulations. Bayesian equivalence results are intended to characterize the priors that ensure identical posterior inference regarding the risk ratio parameters as derived under case-crossover designs and from time-series analysis. Bayesian equivalence results for case-control studies, relating prospective and retrospective likelihoods appear in several recent papers (Seaman and Richardson (2004); Staicu (2010); Ghosh et al. (2012)). The full likelihood formulation requires less restrictive assumptions than the conditional one, however, it involves a set of nuisance parameters corresponding to each individual or day that grows with sample size. Thus maximum likelihood estimators (MLE) of the risk ratio parameters can be potentially inconsistent. We present a semi-parametric Bayesian approach using a Dirichlet process prior (Ferguson (1973); Antoniak (1974); Müller and Quintana (2004)) to handle the random nuisance parameters in the full likelihood formu-

lation. Section 3.3 presents a simulation study where we evaluate the performance of both conditional and full likelihood approaches under two common referent time selection strategies: TSD ('localizable') and SBD ('non-localizable'). We study frequentist properties such as bias and mean-squared error (MSE) of the proposed methods. Our numerical results indicate that Bayesian analysis based on the full likelihood has advantages in relaxing certain model assumptions and reducing bias, but both Bayesian and frequentist inference based on conditional likelihood are fairly robust with respect to design choices and model assumptions. Section 3.4 demonstrates the proposed methods through a study examining the association between acute asthma risk and ambient air pollutant concentrations. We discussed how to use information from published studies through formulation of an informative prior in the context of the example.

We would like to highlight the two fundamentally novel aspects of our study. The present study is the first to consider Bayesian equivalence results between case-crossover and time-series analysis. The proposal to use a full likelihood and use semi-parametric Bayes technique to make the estimation of the nuisance parameters feasible is also completely new in the case-crossover context. The numerical comparison of all proposed Bayesian methods with frequentist alternatives is an added asset of the chapter.

3.2 Methods

3.2.1 Case-crossover and time-series: disease risk model

Suppose that a population of N+M initially disease-free individuals are being followed forward in time. Let time point t (follow-up day, say) stand for the time interval $[t, t+1)$ throughout this chapter. Let Y_{it} be the binary indicator whether subject i has the disease occurring at time t ($Y_{it} = 1$ if yes; $Y_{it} = 0$ if no). Let N and T be the total number of cases at the end of the follow-up period and the total number of the discrete set of time points t

respectively, and let, without loss of generality, the first N of the N+M individuals denote the cases. We start with the risk ratio model using similar notation as in Lu and Zeger (2007), where the risk of an event for individual i at time t is assumed as

$$P\big(Y_{it} = 1 \mid \boldsymbol{X}_{it}\big) = \frac{\lambda_{0it} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})}{1 + \lambda_{0it} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})}. \tag{3.1}$$

$\boldsymbol{X}_{it} = (X_{it1}, ..., X_{itp})^\top$ is the p-dimensional exposure variable for individual i at time t, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ is the common set of log risk ratio parameters. Each individual i is assumed to have his/her own baseline risk λ_{0it}.

If the risk of the disease for individual i at time t is small, it will imply the following:

Assumption 3.1. $P\big(Y_{it} = 1 \mid \boldsymbol{X}_{it}\big) = \lambda_{0it} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it}) / \{1 + \lambda_{0it} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})\} \approx \lambda_{0it} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})$.

For the case-crossover analysis, we do not require assumption 3.1 and can proceed with the likelihood governed by model (3.1); while for the time-series analysis, assumption 3.1 is required.

A traditional conditional likelihood approach

The log risk ratio parameter $\boldsymbol{\beta}$ in model (3.1) can be ascertained under a case-crossover design. We consider the situation where the diseased individuals can have multiple occurrences of events during the whole follow-up period, but the referent windows corresponding to these multiple events from the same individual must not overlap. We ignore the within-individual correlation among the multiple occurrences of events, and treat these multiple events as independent 'cases'. Without loss of generality, we still denote N as the number of 'cases'.

Let t_i and W_i be the event time and referent window for case i. The following assumption on the baseline risk is often made under a case-crossover design, which is natural if the length of W_i is short (typically a month, as in Janes et al. (2005a)).

Assumption 3.2. For each case i with event time t_i, the baseline risk λ_{0it} is constant within the referent window W_i, i.e., $\lambda_{0it} = \lambda_{0it_i}$, for any $t \in W_i$, $i = 1, ..., N$.

Given the exposure $\boldsymbol{X} = (\boldsymbol{X}_{it})_{N\times T}$, the referent windows W_i's and that $\sum_{t\in W_i} Y_{it} = 1$, the conditional likelihood corresponding to a case-crossover design can be derived as

$$
\begin{aligned}
L_{cc}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \prod_{i=1}^{N} P\big(Y_{it_i} = 1, Y_{is} = 0, \forall s \neq t_i \mid \boldsymbol{X}, W_i, \sum_{t\in W_i} Y_{it} = 1\big) \\
&= \prod_{i=1}^{N} \frac{P\big(Y_{it_i} = 1, Y_{is} = 0, \forall s \neq t_i \mid \boldsymbol{X}, W_i\big)}{\sum_{t\in W_i} P\big(Y_{it} = 1, Y_{is} = 0, \forall s \neq t \mid \boldsymbol{X}, W_i\big)} \\
&= \prod_{i=1}^{N} \frac{\lambda_{0it_i} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it_i})}{\sum_{t\in W_i} \lambda_{0it} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})} = \prod_{i=1}^{N} \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it_i})}{\sum_{t\in W_i} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})} = L_{cc}(\boldsymbol{\beta}).
\end{aligned} \tag{3.2}
$$

In (3.2), the third equality holds under a 'localizable' design (Lumley and Levy (2000), Janes et al. (2005b)), under which an unbiased estimate of $\boldsymbol{\beta}$ can be obtained using a CLR. Under assumption 3.2, the fourth equality holds. The nuisance parameter λ_{0it_i} was eliminated by conditioning on the sufficient statistic $\sum_{t\in W_i} Y_{it}$.

The case-crossover design is similar to a matched case-control design in the sense that the exposure at the event time of each case is compared to exposures at all referent times for the same case, i.e., a matched set of exposures corresponding to each W_i. Lumley and Levy (2000) discussed the differences between the two designs, such as the dependency of exposures between and within stratum. Due to these dependencies, they showed that (3.2) can be treated as a conditional likelihood of a matched case-control study only under 'localizable' referent window. They also showed that, with a 'non-localizable' referent window such as a SBD, t_i and W_i are simple functions of each other (t_i is the mid-point of W_i), and $P\big(Y_{it} = 1, Y_{is} = 0, \forall s \neq t \mid \boldsymbol{X}, W_i\big) = I(t = t_i)$ is deterministic. Thus, $L_{cc}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = 1$, i.e., uninformative. The estimating equation for $\boldsymbol{\beta}$ corresponding to $L_{cc}(\boldsymbol{\beta})$ is biased under 'non-localizable' designs, and 'overlap bias' is incurred if $\boldsymbol{\beta}$ is naively estimated using $L_{cc}(\boldsymbol{\beta})$ (Lumley and Levy (2000)).

The case-crossover design is commonly used in ecological studies concerning issues such as effect of climate change and air pollution on human health, where personal exposure is often not assessed at an individual level. For example, ambient air pollutant concentrations are usually measured from monitoring sites representing the exposure of the nearby population.

Assumption 3.3. The study population has experienced shared exposure at each time t such that $\boldsymbol{X}_{it} = \boldsymbol{X}_t$, for $i = 1, ..., N$+M.

Under assumption 3.3, the conditional likelihood in (3.2) can be rearranged in terms of the number of events (e.g. daily mortality) at each follow-up time t (e.g. day) and expressed as

$$L_{cc}(\boldsymbol{\beta}) = \prod_{t=1}^{T} \Big\{ \prod_{i:\, Y_{it}=1} \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})}{\sum_{s\in W_i} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{is})} \Big\} = \prod_{t=1}^{T} \Big\{ \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{t})}{\sum_{s\in W(t)} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{s})} \Big\}^{Y_t}, \tag{3.3}$$

where $W(t)$ is the referent window containing t as the event time, $Y_t = \sum_{i=1}^{N} Y_{it}$ is the count of events at time t. Assumption 3.3 can be relaxed and individual exposure values can be accommodated as in our earlier formulation (3.1) or (3.2).

A full likelihood approach

We propose an alternative full likelihood formulation of model (3.1). From (3.1), before enforcing either assumption 3.2 or 3.3, the full likelihood of a case-crossover design is expressed in terms of individual level exposure and baseline risk as

$$\begin{aligned} L^{NT}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \prod_{i=1}^{N} P\big(Y_{it_i} = 1, Y_{is} = 0, \forall s \neq t_i,\ s \in W_i \mid \boldsymbol{X}\big) \\ &= \prod_{i=1}^{N} \frac{\lambda_{0it_i} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it_i})}{\prod_{s\in W_i} \{1 + \lambda_{0is} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{is})\}}, \end{aligned} \tag{3.4}$$

which allows for a completely general form of λ_{0it}. We refer to $L^{NT}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ in (3.4) as full likelihood with individual and day level intercepts. If the baseline risk for individual i

does not change in its referent window and depends only on the event time t_i as in assumption 3.2, then we can write (3.4) as $L^N_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda}) = \prod_{i=1}^{N} \Big[\lambda_{0it_i} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it_i}) / \prod_{s \in W_i} \{1 + \lambda_{0it_i} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{is})\}\Big]$. We refer to $L^N_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ as full likelihood with individual level intercepts. In this case, the full likelihood $L^N_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ under a case-crossover design is exactly analogous to deriving the full likelihood of a matched case-control study under a stratified logistic regression model $\text{logit}\, P(Y_{is} = 1 \mid \boldsymbol{X}_{is}) = \log(\lambda_{0it_i}) + \boldsymbol{\beta}^\top \boldsymbol{X}_{is}, s \in W_i$, $i = 1, ..., N$. Under both assumptions 2 and 3, one can alternatively translate the likelihood in terms of common nuisance parameters $\log(\lambda_{0it_i}) = \nu_{t_i}$ for all cases i that have event time t_i. Then $L^N_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$ can be aggregated as

$$L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu}) = \prod_{t=1}^{T} \Big[\frac{\exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t)}{\prod_{s \in W(t)} \{1 + \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_s)\}}\Big]^{Y_t}, \tag{3.5}$$

where $\boldsymbol{\nu} = (\nu_1, .., \nu_T)$. We refer to $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ as full likelihood with day level intercepts.

Time-series Analysis

The log risk ratio parameter $\boldsymbol{\beta}$ in model (3.1) can also be estimated using an alternative time-series analysis. The expected number of events at time t can be expressed as the sum of the individual level probabilities (3.1) over the population. Under both assumptions 1 and 3, $\log(E(Y_t)) = \log(\sum_{i=1}^{N+M} E(Y_{it})) = \boldsymbol{\beta}^\top \boldsymbol{X}_t + \log(\sum_{i=1}^{N+M} \lambda_{0it})$. For exposures varying at an individual level $\boldsymbol{X}_{it}$, generally one can not aggregate the risk $\lambda_{0it} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_{it})$ as above to obtain a standard time-series structure involving the day level counts and exposures. Consider $\log(\sum_{i=1}^{N+M} \lambda_{0it})$ as a function of time only, say S_t. Then $\boldsymbol{\beta}$ can be estimated through the log-linear Poisson model

$$\log(E(Y_t)) = \boldsymbol{\beta}^\top \boldsymbol{X}_t + S_t, \tag{3.6}$$

where S_t is typically modeled as parametric (e.g., season) and/or non-parametric (e.g., natural spline of time) terms. The likelihood corresponding to the log-linear model in

(3.6) is given by

$$L_{ll}(\boldsymbol{\beta}, S_t) \propto \prod_{t=1}^{T} \{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S_t)\}^{Y_t} \exp\{-\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S_t)\}. \tag{3.7}$$

Frequentist equivalence between time-series analysis and case-crossover design using conditional likelihood

The two estimating equations for $\boldsymbol{\beta}$ corresponding to (3.3) and (3.6) are $U_{cc}(\beta) = \sum_{t=1}^{T} X_t \Big\{ Y_t - \exp(\beta^\top X_t) \sum_{s \in W(t)} Y_s / \sum_{r \in W(s)} \exp(\beta^\top X_r) \Big\}$ and $U_{ll}(\beta) = \sum_{t=1}^{T} X_t \Big\{ Y_t - \exp(\beta^\top X_t + S_t) \Big\}$ respectively. By comparing $U_{cc}(\beta)$ and $U_{ll}(\beta)$, Lu and Zeger (2007) showed that, for a certain choice of window $W(t)$ in (3.3) of a 'localizable' design, there exists a choice of S_t in log-linear model (3.6) such that the two estimating equations provide the same estimate of $\boldsymbol{\beta}$. For example, for a TSD with $W(t)$ representing the time stratum containing time t, if $\widehat{S}_t(\boldsymbol{\beta}) = \log[\{\sum_{s \in W(t)} Y_s\} / \{\sum_{s \in W(t)} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_s)\}]$, then log-linear model (3.6) will provide the same estimate of $\boldsymbol{\beta}$ as (3.3). Note that $\widehat{S}_{t'}(\boldsymbol{\beta}) = \widehat{S}_t(\boldsymbol{\beta})$ for any $t' \in W(t)$, implying S_t is a step function of t with a separate value at each time stratum. As the conditional likelihood is uninformative under a 'non-localizable' design, there is no equivalence between time-series analysis using a log-linear model and case-crossover design.

Frequentist equivalence using full likelihood

Similarly, by comparing the two estimating equations corresponding to $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ and $L_{ll}(\boldsymbol{\beta}, S_t)$, we showed (in Appendix B.2) that, for a certain choice of window $W(t)$ of a 'localizable' design, there exists a choice of S_t in the log-linear model such that the two estimating equations provide the same estimate of $\boldsymbol{\beta}$. Under a TSD, while the conditional likelihood approach or an equivalent log-linear model would only allow the baseline risk to change discontinuously among different time strata, the full likelihood approach does not

require such constraints. However, both the full likelihood method and its equivalent log-linear model encounter difficulty in estimating $\boldsymbol{\beta}$ in the presence of T day level nuisance parameters using maximum likelihood. As an alternative estimation strategy, a random effects Bayesian approach could be used to handle these random nuisance parameters under the full likelihood formulations, which are described in section 3.2.3.

3.2.2 Bayesian equivalence between case-crossover design and time-series model

Bayesian equivalence with conditional likelihood

We focus on the posterior distributions of the log risk ratio parameter $\boldsymbol{\beta}$ derived under case-crossover and time-series analysis. The Bayesian equivalence result for $\boldsymbol{\beta}$ requires that the posterior distribution of $\boldsymbol{\beta}$ derived from $L_{cc}(\boldsymbol{\beta})$ in (3.3) and from $L_{ll}(\boldsymbol{\beta}, S_t)$ in (3.7) are identical under certain forms of S_t and certain prior distributions on $\boldsymbol{\beta}$ and S_t. The validity of using a conditional likelihood as the basis for Bayesian inference has been discussed in previous studies. For example, Rice (2004) and Rice (2008) discussed the equivalence between the use of conditional and marginal likelihoods for matched case-control study. Since the conditional likelihood approach is only valid for a 'localizable' design, we restrict our attention specifically to a TSD in this section.

THEOREM 3.1. Suppose the follow-up time points $t = 1, ..., T$ are divided *a priori* into K disjoint time strata $ts(k)$ under a TSD, $k = 1, ..., K$. If S_t in log-linear model (3.6) is defined as a step function with distinct values of S_t be S'_k on $ts(k)$, $k = 1, ..., K$, and if independent improper priors $\pi(S'_k) \propto 1$ for S'_k and a proper prior $\pi(\boldsymbol{\beta})$ for $\boldsymbol{\beta}$ are used where S'_k and $\boldsymbol{\beta}$ are mutually independent, then the posterior distribution of $\boldsymbol{\beta}$ derived from $L_{cc}(\boldsymbol{\beta})$ in (3.3) is identical to the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L_{ll}(\boldsymbol{\beta}, S_t)$ in (3.7).

Proof of Theorem 3.1 is given in Appendix B.2. We showed that, given the choice of S_t

and prior distribution on S_t and $\boldsymbol{\beta}$ as in Theorem 3.1, the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L_{ll}(\boldsymbol{\beta}, S_t)$ is $\pi(\boldsymbol{\beta} \mid \boldsymbol{X}, \boldsymbol{Y}) \propto \int \cdots \int \pi(\boldsymbol{\beta})\pi(S'_1, ..., S'_K)L_{ll}(\boldsymbol{\beta}, S_t)dS'_1 \cdots dS'_K$ $\propto \pi(\boldsymbol{\beta})L_{cc}(\boldsymbol{\beta})$, i.e., the posterior distribution of $\boldsymbol{\beta}$ derived from $L_{cc}(\boldsymbol{\beta})$.

With shared exposure data across all individuals, the time-series model in (3.6) is more flexible than a case-crossover design using conditional likelihood, in the sense that model (3.6) allows various smoothing functions of time for S_t where one special choice is equivalent to the analysis of data under case-crossover design. Log-linear models can also account for over-dispersion of the Poisson variance that is typically present in air pollution studies, while case-crossover studies can not. In contrast, case-crossover design has the advantage of controlling for personal level confounders, and modeling individual level exposures over time-series models.

Bayesian equivalence with full likelihood

We aim to show the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L^T_{full}(\boldsymbol{\nu}, \boldsymbol{\beta})$ under the shared exposure assumption is identical to that derived from a Poisson likelihood. Let y_{s1t} and y_{s0t} be the numbers of potential event and referent times that equals to t in the s-th time stratum. y_{sdt} is assumed to follow a Poisson distribution $Poisson(\mu_{sdt})$ with mean $\mu_{sdt} = \exp(\phi_{st} + d(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t))$, $d = 0, 1$; $s = 1, ..., T$; $t = 1, ..., T$. The Poisson likelihood with ancillary parameters ϕ_{st} is given by,

$$L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta}) \;=\; \prod_{s=1}^{T}\prod_{t=1}^{T}\prod_{d=0}^{1}\big[\exp\{\phi_{st} + d(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t)\}\big]^{y_{sdt}} \exp\big[-\exp\{\phi_{st} + d(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t)\}\big]$$

With independent improper priors $\pi(\phi_{st}) \propto 1$ and proper prior on $\boldsymbol{\nu}$ and $\boldsymbol{\beta}$, the joint posterior distribution of $(\boldsymbol{\nu}, \boldsymbol{\beta})$ derived (in Appendix B.2) from $L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta})$ is

$$\pi(\boldsymbol{\nu}, \boldsymbol{\beta} \mid \boldsymbol{X}, \boldsymbol{Y}) \propto \int \pi(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta})L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta})d\boldsymbol{\phi} \propto \pi(\boldsymbol{\nu}, \boldsymbol{\beta})L^T_{full}(\boldsymbol{\nu}, \boldsymbol{\beta}).$$

So the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta})$ and from $L^T_{full}(\boldsymbol{\nu}, \boldsymbol{\beta})$ are the same. This method is inspired by the Multinomial-Poisson transformation (Baker 1994) and its Bayesian counterpart (Seaman and Richardson (2004); Ghosh et al. (2006)). Though this proves theoretical Bayesian equivalence between using a case-crossover full likelihood and a Poisson likelihood, the Poisson model has a large number of nuisance parameters $\boldsymbol{\phi}$. Moreover, the interpretation of the artificially constructed Poisson model is practically not very meaningful. Thus, we focus on full likelihood based methods with more flexible semi-parametric prior distributions on the intercepts in the following section, instead of an equivalent time-series formulation.

3.2.3 Bayesian inference

Traditionally, CLR models were routinely used for frequentist inference on $\boldsymbol{\beta}$ under case-crossover designs. A naive approach would be to also use the conditional likelihood $L_{cc}(\boldsymbol{\beta})$ as the basis for Bayesian inference, where prior specification on only $\boldsymbol{\beta}$ is needed. The posterior distribution of $\boldsymbol{\beta}$ is not a standard distribution, but posterior draws could be generated by using a Gibbs sampler (Appendix B.2).

For the full likelihood approach, though the number of nuisance parameters grows with the sample size, a random effects model can be used to reduce the problem to estimating the parameters corresponding to the random effects distribution. For example, the stratified logistic regression model with likelihood $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ in (3.5) can be readily fitted through a generalized linear mixed model with $\nu_t \overset{iid}{\sim} N(\mu_\nu, \sigma^2_\nu)$. Methods such as penalized pseudo-likelihood (Breslow and Clayton 1993) can be used for inference in such models, which is available in standard statistical software.

The misspecification of the random effects distribution could lead to potential bias in the estimation of $\boldsymbol{\beta}$ (Mukherjee et al. 2009). To avoid assuming a parametric nor-

mal distribution on the nuisance parameters, we consider a more robust semi-parametric Bayesian approach that allows the random effects to have a nonparametric distribution. To this end, we use the Dirichlet process prior (Ferguson (1973); Antoniak (1974); Müller and Quintana (2004)) to handle the random intercepts. In particular, for example with $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ in (3.5), we assume $\nu_t \mid G \overset{iid}{\sim} G$, where G is a random distribution generated from a Dirichlet process with concentration parameter α and base distribution G_0, i.e., $G \mid \alpha, G_0 \sim DP(\alpha, G_0)$. Let $\boldsymbol{\nu}_{-t} = (\nu_1, ..., \nu_{t-1}, \nu_{t+1}, ..., \nu_T)$, for $t = 1, ..., T$. The joint prior distribution $\pi(\boldsymbol{\nu})$ can be represented in terms of leave-one-out conditional distributions as $\nu_t \mid \boldsymbol{\nu}_{-t} \sim \frac{\alpha}{T-1+\alpha} G_0 + \frac{1}{T-1+\alpha} \sum_{s=1,\, s\neq t}^{T} I_{\nu_s}(\cdot)$ (Blackwell and MacQueen 1973). Thus, $(\nu_1, ..., \nu_T)$ will be adaptively reduced to fewer distinct clusters with positive probability. As $\alpha \to \infty$, the Dirichlet process model reduces to specifying a parametric model $\nu_t \overset{iid}{\sim} G_0$; whereas $\alpha \to 0$ implies a parametric model with a common stratum effect, namely $\nu_t = \nu^*$ for $t = 1, ..., T$, where $\nu^* \sim G_0$.

To complete the hierarchy, independent hyperpriors are considered as follows: $\alpha \mid a_0, b_0 \sim Gamma(a_0, b_0)$, $G_0 \sim N(\mu, \sigma^2)$, $\mu \mid \mu_0, \sigma_0 \sim N(\mu_0, \sigma_0^2)$, $\sigma^{-2} \mid c_0, d_0 \sim Gamma(c_0, d_0)$. We consider mutually independent normal priors $\boldsymbol{\beta} \sim N(\mu_\beta, \sigma^2_\beta \boldsymbol{I}_p)$. The posterior distributions of $\boldsymbol{\nu}$ and $\boldsymbol{\beta}$ can be obtained by using a Metropolis-Hastings within Gibbs algorithm as described in Neal (2000) and Sinha et al. (2004). The details are presented in Appendix B.2. Similarly, for the individual specific stratum effects in $L^N_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$, we assume $\log(\lambda_{0it_i}) \mid G \overset{iid}{\sim} G$, for $i = 1, 2, ..., N$. For $L^{NT}_{full}(\boldsymbol{\beta}, \boldsymbol{\lambda})$, we only consider a special case by assuming a multiplicative structure on the nuisance parameters: $\lambda_{0it} = \lambda_{0i} \exp(\omega_t)$, where λ_{0i} is a constant frailty for person i and ω_t is the time varying effect on the risk. We model $\log(\lambda_{0i})$ and ω_t through random distributions generated from two independent Dirichlet processes.

3.3 A Simulation Study

Simulation scenarios

In our simulation, we used $\lambda_{0it} = \lambda_{0i}\exp(\omega_t)$ as the form of the true baseline risk. Under both assumptions 1 and 3, we have $\log(E(Y_t)) = \boldsymbol{\beta}^\top \boldsymbol{X}_t + \log(\sum_{i=1}^{N+M}\lambda_{0it}) = \boldsymbol{\beta}^\top \boldsymbol{X}_t + \omega_t + \eta$, where $\eta = \log(\sum_{i=1}^{N+M}\lambda_{0i})$. We generated the number of events per day from a Poisson model

$$Y_t \sim Poisson(\mu_t), \quad \text{where } \mu_t = \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t + \omega_t + \eta). \tag{3.8}$$

We considered various simulation scenarios with different choices of time effects ω_t on the baseline risk, true effect sizes β^*, and exposure series X_t. Without loss of generality, we considered $\boldsymbol{\beta}$ and $\boldsymbol{X}_t$ to be univariate in our simulation study. We convert the time-series data in the form of individual event referent times according to a given case-crossover design.

Temporal trends on the baseline risk: In order to examine whether the full likelihood method under a case-crossover design is more robust to various baseline risk specifications than the conditional likelihood method, we considered three forms of time-varying effect ω_t involved in the baseline risk. In particular, B1: $\omega_t = \omega$ that satisfies assumption 3.2; B2: $\omega_t = c(1 - 0.001t)[1 + 0.5cos(2\pi t/365)]$ that combines seasonal and long-term decreasing trends, where c is a positive scaling factor; B3: B3 is a mixture of B2 (with probability 0.9) and random spikes (with probability 0.1), where the spikes follow a uniform distribution $U(2c, 4c)$. Note that B2 and B3 both violate assumption 3.2.

Effect sizes: We considered two typical true effect sizes $\beta^* = 0.1$ (a risk ratio of 1.1, e.g. reflecting the effect of 10 $\mu g\, m^{-3}$ increase of $PM_{2.5}$ (particulate matter less than 2.5 micrometers in diameter) on the risk of acute asthma (Li et al. 2011) or 10 degree (F) change of temperature on mortality risk (Basu et al. 2005)) and $\beta^* = 1$ (a risk ratio of 2.7,

e.g. reflecting the effect of medication use on preventing elderly falls in case-crossover intervention trials (Luo and Sorock 2008)).

Exposure series: We simulated exposure series X_t over a 3-year (T=1096) period under two settings. E1: X_t has auto-correlation structure $AR(1)$ ($\rho = 0.6$); E2: X_t has long-term decreasing trend plus seasonal and day of week effects, with the same auto-correlation structure as in E1. We generated E1 and E2 to have the same marginal distributions.

Likelihoods/Methods: We would like to compare time-series analysis using log-linear models with the analysis under case-crossover designs. The methods can broadly be divided into three classes in terms of likelihoods we considered. In particular, M1: log-linear models, adjusted for the true temporal trend ω_t as offset (this is the closest to the true generating model) or adjusted for a natural spline term on time t; M2: conditional likelihood approach under a case-crossover design, using both frequentist and Bayesian treatment; M3: full likelihood approach under a case-crossover design, with the random intercepts handled by a Dirichlet process. Within case-crossover analysis, we also compared the two commonly used referent time selection strategies, TSD with SBD, to possibly quantify the 'overlap bias'.

Prior choices: Within the Bayesian alternatives we implemented two choices of prior distributions on β: non-informative and informative. For non-informative prior, we used a vague prior $\beta \sim N(0, 10^2)$; for informative prior, we considered $\beta \sim N(\mu_\beta, \sigma_\beta^2)$ with μ_β and σ_β potentially elicited from historical data. We described the details of incorporating historical data to construct informative priors under a concrete data example framework in section 3.4. We used $\beta \sim N(0.08, 0.03^2)$ when $\beta^* = 0.1$, and $\beta \sim N(0.8, 0.2^2)$ when $\beta^* = 1$ in our simulations. Justification for these prior choices on μ_β and σ_β as well as on the full set of parameters are provided in Appendix B.2.

The simulation was repeated 1000 times under each scenario. We summarized the

results in terms of relative bias ($RB = (\frac{1}{1000}\sum_{i=1}^{1000}\hat{\beta}_i - \beta^*)/\beta^* \times 100\%$) and MSE ($MSE = \frac{1}{1000}\sum_{i=1}^{1000}(\hat{\beta}_i - \beta^*)^2$) corresponding to the log risk ratio parameter β. Tables 3.1 and 3.2 present results corresponding to 6 baseline$\times$exposure (3×2) settings, for $\beta^* = 0.1$ and 1 respectively. As the individual level model $L^{NT}_{full}(\beta, \boldsymbol{\lambda})$ and $L^{N}_{full}(\beta, \boldsymbol{\lambda})$ are computationally intensive with large N ($N \approx 20,000$ for $\beta^* = 0.1$ in our simulations), we only considered $L^{T}_{full}(\beta, \boldsymbol{\nu})$ under $\beta^* = 0.1$. We considered all three versions of full likelihoods under $\beta^* = 1$ with $N \approx 1000$.

Simulation results

Likelihoods/Methods: We present the estimates from the log-linear model using offset terms to be the true values of ω_t as a reference benchmark, against which each of our methods is compared. In practice, while carrying out a time-series analysis, one will not know the true time effect terms and will use a flexible nonparametric spline term (Dominici et al. (2002), Dominici et al. (2003)). The log-linear model adjusted for a natural cubic spline term of time with 7 degrees of freedom per year approximates the true model quite accurately. The two log-linear models both have smaller bias and MSE than the case-crossover designs, especially under B2 or B3. Note that the log-linear models fitted here were not chosen to be the equivalent models to case-crossover designs.

(a) Design effect: In comparing the two designs, we found that the TSD generally has smaller bias and MSE than the SBD. Under B1, the only source of bias is the 'overlap bias' of a SBD. We observed up to 5% difference in bias between TSD and SBD (Tables 3.1 and 3.2). However, the magnitude and direction of the 'overlap bias' depend on the particular exposure series and effect size, as previously noted in Janes et al. (2005a).

(b) Conditional versus full likelihood formulation (with non-informative priors): The conditional likelihood $L_{cc}(\beta)$ as well as the full likelihood $L^{N}_{full}(\beta, \boldsymbol{\lambda})$ and $L^{T}_{full}(\beta, \boldsymbol{\nu})$

require assumption 3.2 of constancy of ω_t in each referent window and only allows the risk to change discontinuously across referent windows. The most general form of the full likelihood $L^{NT}_{full}(\beta, \boldsymbol{\lambda})$ does not require this assumption. As for bias, we note that the bias due to violation of assumption 3.2 (under B2 and B3) is typically very small ($< 1\%$) when $\beta^* = 0.1$ (Table 3.1), and up to 3% when $\beta^* = 1$ (Table 3.2). However, full likelihood methods have greater MSE than conditional counterparts under non-informative priors. This is expected as another level of hierarchy was added to model the uncertainty in the random nuisance parameters.

Prior sensitivity: Both conditional and full likelihood methods show substantial reduction in MSE when informative priors on β are used as shown in Tables 3.1 and 3.2, which also lead to substantial shrinkage towards the prior mean. Thus, given the context of the study and prior information, Bayesian methods that utilize informative priors can potentially have advantage over their frequentist counterparts in terms of MSE.

Remark 3.1: Note that the full likelihood has additional analytic flexibility to handle individual level data and incorporate interaction terms that account for subject level covariates. We generated time-series data with shared exposure for the present simulation study for illustration purposes, leading to best performance by the log-linear models. If we had generated individual level data with personal factors, the time-series analysis would have been more susceptible to residual bias from ignoring personal level confounders.

3.4 Application: Data analysis for the DAMAT Study

We illustrate the proposed methods through the 'Detroit Asthma Morbidity, Air Quality and Traffic' (DAMAT) study originally analyzed by Li et al. (2011). One primary goal of the study was to examine the association between acute asthma risk and ambient air pollutant levels, especially $PM_{2.5}$, for the pediatric (children 2-18 years) Medicaid population

in Detroit, Michigan, during the 2004-2006 study period ($T = 1096$ days). Daily counts of asthma events, including emergency department visits and hospitalizations, were used as the outcome series Y_t, for $t = 1, ..., 1096$. Figure 3.1 shows the smoothed trend of Y_t indicating a strong seasonal pattern, with the highest frequency during fall, and the lowest during summer. A total of 12,933 asthma events were observed during the 1096 days, representing an average rate of 11.8 events per day. Daily $PM_{2.5}$ concentration was computed as the average concentration across the air quality monitoring sites in the Detroit area. Daily $PM_{2.5}$ data also show a strong seasonal pattern with a mean level of 15.0 $\mu g\, m^{-3}$. Daily meteorological variables, including temperature (TP) and relative humidity (RH), were obtained similarly. To account for other temporal trends that were not controlled by the case-crossover design, a natural quadratic spline term (denoted by $ns(\cdot)$) was used on the TP variable.

Likelihoods: Under the case-crossover design, the conditional likelihood is given by

$$L_{cc}(\boldsymbol{\beta}) = \prod_{t=1}^{1096} \Big[\frac{\exp\{\beta_{PM_{2.5}} \mathrm{PM}_{2.5,\,t} + \beta_{RH} \mathrm{RH}_t + ns(\mathrm{TP}_t)\}}{\sum_{s \in W(t)} \exp\{\beta_{PM_{2.5}} \mathrm{PM}_{2.5,\,s} + \beta_{RH} \mathrm{RH}_s + ns(\mathrm{TP}_s)\}} \Big]^{Y_t}, \tag{3.9}$$

where we used the 5-day moving average of $PM_{2.5}$ concentration prior to the asthma events on day t as $\mathrm{PM}_{2.5,\,t}$. The Medicaid data was also analyzed using the equivalent log-linear model of (3.9). In particular, for a TSD with time stratum as the same day of the week in the same calender month, we compare it with the following equivalent log-linear model

$$\log(E(Y_t)) = \beta_0 + \beta_{PM_{2.5}} \mathrm{PM}_{2.5,\,t} + \beta_{RH} \mathrm{RH}_t + ns(\mathrm{TP}_t) + S_t, \tag{3.10}$$

where S_t represents all possible combinations among the three factors of day of the week, month and year, having a total of $7 \times 12 \times 3 = 252$ levels. We refer to (3.10) as the time-stratified log-linear (TSLL) model. For inference on $\boldsymbol{\beta}$, models (3.9) and (3.10) are equivalent under both the frequentist and Bayesian framework (with prior specification as described in Theorem 3.1). Similarly, a case-crossover SBD using referent times as 7 and 14

days before and after the event day was compared to the corresponding symmetric bidirectional log-linear (SBLL) model with $S_t = \log[\sum_{s=t,\, t\pm 7,\, t\pm 14}\{Y_s / \sum_{r=s,\, s\pm 7,\, s\pm 14} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_r)\}]$. Joint estimation of $\boldsymbol{\beta}$ and S_t was performed iteratively as S_t potentially depends on $\boldsymbol{\beta}$. The comparison between SBD and SBLL is pertinent only under the frequentist framework.

Prior choices: We used $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ under the full likelihood approach. We first considered the random effects model assuming $\nu_t \overset{iid}{\sim} N(0, 10^2)$ without further prior specification on $\boldsymbol{\beta}$, and proceeded with the marginal likelihood to estimate $\boldsymbol{\beta}$. Then we considered the full Bayesian treatment using Dirichlet process prior $\nu_t \mid G \overset{iid}{\sim} G$ where $G \mid \alpha, G_0 \sim DP(\alpha, G_0)$. $\alpha \sim Gamma(0.5, 0.1)$, $G_0 \sim N(\mu, \sigma^2)$, $\mu \sim N(0, 10)$ and $\sigma^{-2} \sim Gamma(4, 1)$ were used as the base prior setting in our data example. As part of our sensitivity analysis, we varied the priors on α across four $Gamma$ distributions, and two extreme cases when $\alpha \to 0$ (corresponding to $\nu_t = \nu^*$ for $t = 1, ..., T$, where $\nu^* \sim G_0$) and $\alpha \to \infty$ (corresponding to $\nu_t \overset{iid}{\sim} G_0$). More details were provided in Figure 3.2(c).

We considered both informative and non-informative priors on $\beta_{PM_{2.5}}$. For noninformative prior, we used a vague prior $\beta_{PM_{2.5}} \sim N(0, 10^2)$. For informative prior, we considered an ad-hoc way of eliciting prior information from published results. From a recent review (Li et al. 2011), we *a priori* postulated that the asthma-$PM_{2.5}$ association is in general modest with a risk ratio ranging in (1.01-1.09) for 10 $\mu g\, m^{-3}$ increase in $PM_{2.5}$. Assuming $\beta_{PM_{2.5}} \sim N(\mu_\beta, \sigma^2_\beta)$, if we believed that the 95% confidence interval (CI) for $\exp(\beta_{PM_{2.5}})$ is (1.01,1.09), the approximate values for μ_β and σ_β can be obtained as $\mu_\beta = [\log(1.09) + \log(1.01)]/2 = 0.05$ and $\sigma_\beta = [\log(1.09) - \log(1.01)]/4 = 0.02$. Then our informative prior was chosen as $N(0.05, 0.02^2)$. There is no general consensus on the best way to elicit a subjective prior though this topic has been studied vastly (Dey and Liu 2007).

While non-informative or non-subjective priors are often quite adequate as default priors for many Bayesian analyses, prior elicitation when possible can lead to more meaningful results from the data. This is especially true in the presence of historical data, available from series of past studies, e.g., past studies relating asthma-$PM_{2.5}$ associations in our context. These elicited priors based on historical data, when proper, have an operational advantage over non-subjective improper priors. Proper priors are required for computing Bayes factors and posterior model probabilities.

Ibrahim and Chen (2000) proposed a particular approach towards the development of priors based on historical data. They referred to these priors as 'power priors'. Specifically, if D_0 denotes the historical data from a previous study, the power prior for β is defined as $\pi(\beta|D_0, a_0, c_0) \propto L^{a_0}(\beta|D_0)\pi_0(\beta|c_0)$. Here $\pi_0(\beta|c_0)$ is the initial prior before the data D_0 were observed and c_0 is a specified hyperparameter. The parameter $a_0 \in [0, 1]$ is a scalar parameter which controls the influence of the historical data on the current data. In particular, $a_0 = 1$ corresponds to the past posterior which has become the present prior. On the other hand, $a_0 = 0$ corresponds to a prior specification which ignores completely the historical data. While one can do the analysis by simply assigning also a prior on a_0, we will pursue our analysis both when a_0 is fixed or random, assigning *Beta* priors to a_0 in the latter case. If $\pi_0(\beta|c_0)$ is proper, then the power prior $\pi(\beta|D_0, a_0, c_0)$ is guaranteed to be proper; further, $\pi(\beta|D_0, a_0, c_0)$ can be proper under certain regression settings even if $\pi_0(\beta|c_0)$ is an improper uniform prior (Ibrahim and Chen 2000). We used a vague initial prior $N(0, 10^2)$ for $\pi_0(\beta|c_0)$, such that the prior $\pi(\beta|D_0, a_0, c_0)$ is a handy proper prior. $L(\beta|D_0)$ was constructed based on several published studies of asthma-$PM_{2.5}$ associations reviewed in (Li et al. 2011). Further details regarding the construction of $L(\beta|D_0)$ are given in Appendix B.2.

Results: The results are shown in Table 3.3 and Figure 3.2. In general, evidence of signif-

icant increases in acute asthma risk was found with 10 $\mu g\, m^{-3}$ increase in $PM_{2.5}$ concentrations leading to a risk ratio ranging from 1.02 to 1.06 across different methods.

(a) Design effect: Comparing TSD with SBD, $\hat{\beta}_{PM_{2.5}}$ estimated under a TSD (ranging from 1.05 to 1.06 in Table 3.3) are larger than those estimated under a SBD (from 1.02 to 1.04). The overall pattern of the attenuated effects under the SBD is probably due to the choice of the window and potential 'overlap bias', though this direction does not hold in general as noted in our simulation study.

Comparing the case-crossover TSD with corresponding TSLL model (in Table 3.3), we noted that they provided identical numerical results for $\hat{\beta}_{PM_{2.5}}$ under both frequentist and Bayesian framework (except possible Monte Carlo errors), indicating the numerical validity of our equivalence results. Frequentist equivalence results also appear to hold numerically for the case-crossover SBD and the corresponding time-series SBLL.

(b) Conditional versus full likelihood: Under a case-crossover TSD, full likelihood methods provided slightly stronger effects (risk ratio ranging from 1.05 to 1.06 in Table 3.3) than those derived using conditional likelihood (from 1.04 to 1.05). As noted in our simulation study, the violation of the constant baseline risk assumption within each window probably led to this difference. We also fit a log-linear model adjusted for a natural cubic spline term of time with 7 degrees of freedom per year, which shows an estimate of 1.055 (95% CI: (1.027, 1.084)) that is similar to the results using $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$. Under SBD, there is no substantial difference of using full versus conditional likelihood.

Prior sensitivity: **(a)** Priors on $\beta_{PM_{2.5}}$: When a vague prior $\beta_{PM_{2.5}} \sim N(0, 10^2)$ is used, the Bayesian approaches yielded results that are quite similar to maximum likelihood-based inferences (Table 3.3). It is reassuring that with modest to large sample sizes, we observed similar results from Bayesian and frequentist methods. With a smaller sample size using only one-year data ($T = 365$), even the use of non-informative prior increased

the precision as compared to the frequentist methods. In Table 3.3, the use of informative priors (including the ad-hoc prior 2 and the four power priors) increased the precision as compared to the results under the vague prior. Given one historical study having relatively larger effect (Appendix B.2), the use of power priors 2 ($a_0 = 1$) and 4 ($a_0 \sim Beta(50, 1)$ with prior mean ≈ 1) provided stronger effects than those under power priors 1 ($a_0 = 0.5$) and 3 ($a_0 \sim Beta(20, 20)$ with prior mean 0.5), because power priors 2 and 4 put more weight on $L(\beta|D_0)$. The use of another layer of uncertainty on a_0 (power priors 3 and 4) creates a heavier tail for the marginal power prior distribution of $\beta_{PM_{2.5}}$ than that using a fixed a_0 (power priors 1 and 2), and thus provides wider HPD credible intervals.

(b) Priors on $\boldsymbol{\nu}$: Figure 3.2(c) shows the posterior distributions of $\beta_{PM_{2.5}}$ derived under the 6 different prior settings on α, where the posterior distributions of $\beta_{PM_{2.5}}$ remain robust. We observed that the random effects model and the Dirichlet process prior model provided very similar results in Table 3.3. Although the prior support allows the number of clusters ranging from 1 to 252, we observed only 1 cluster under 4 out of the 6 prior settings (shown in Appendix A.2 Table 1). The results suggested that a parametric constant random intercept model was adequate for this data set.

(c) Priors on S_t: For TSLL (3.10), we considered a sensitivity analysis of prior on S_t instead of the flat prior ($\propto 1$) indicated in Theorem 3.1. In particular, a Dirichlet process prior as well as an i.i.d. normal prior $N(0, 10^2)$ on S'_k was also used. Figure 3.2(d) shows that the posterior distribution of $\beta_{PM_{2.5}}$ remained very similar for all these priors on S'_k. So there is evidence that the results in Theorem 3.1 are robust with respect to prior specification on S_t.

3.5 Discussion

The chapter presents two novel ideas in the context of case-crossover studies, and it is the first treatment of the problem in a Bayesian domain. The first contribution is to study equivalence properties in terms of obtaining identical posterior inference under case-crossover and time-series analysis. The second and more important contribution is to propose different forms of full likelihood and strategies for flexible semi-parametric Bayesian estimation and inference under such likelihoods. Our numerical example and simulation studies illustrate that the Bayesian specification has advantages in terms of robustness to model misspecification on the baseline risks and efficiency advantages if an informative prior is used on the risk ratio parameter. A major potential advantage for using the full likelihood could be to include individual level data. This formulation makes it possible to test for evidence of effect modification of exposure effect by individual level factors, an analysis that is not feasible under a conditional likelihood formulation.

This work leads to many other potential extensions where a Bayesian analysis may have attractive features under a case-crossover design. For example, extensions to distributed lag linear/non-linear models (Welty et al. (2009); Gasparrini et al. (2010)), hierarchical models for meta-analysis (e.g., Dominici et al. (2000), Dominici et al. (2002)), and recurrent events (Luo and Sorock 2008) are natural directions to pursue.

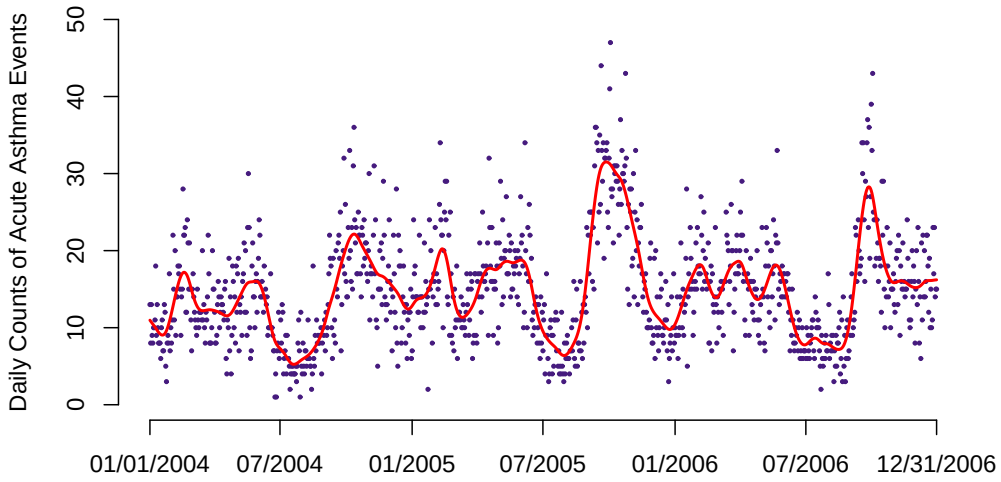
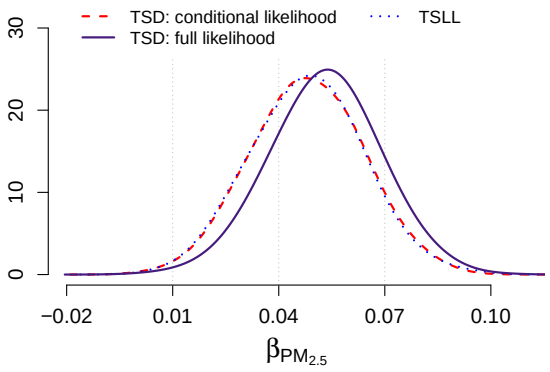
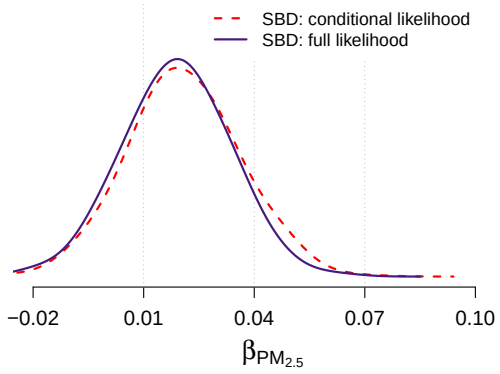


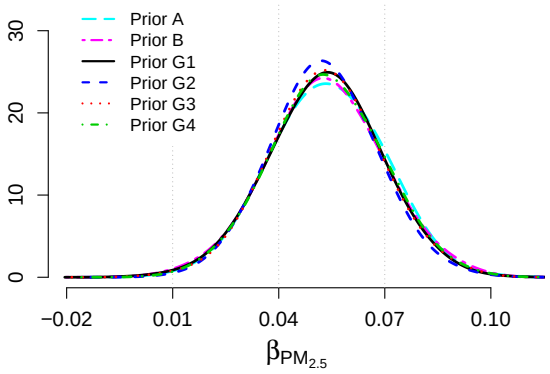
Figure 3.1: Temporal trend of daily counts of acute asthma events (shown as points) for the pediatric Medicaid population in Detroit, Michigan, 2004-2006, as obtained in the DAMAT study. The overlaying smooth curve is created by using locally estimated scatter-plot smoother.



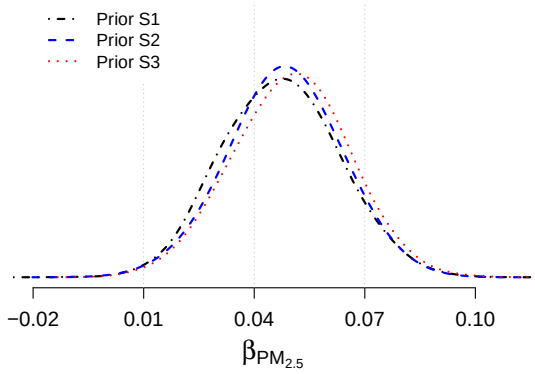
(a) Posterior densities using likelihood under TSD



(b) Posterior densities using likelihood under SBD



(c) TSD using $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ and 6 different DP priors



(d) TSLL using 3 different priors on S_t

Figure 3.2: Posterior density plots for the log risk ratio parameter $\beta_{PM_{2.5}}$ corresponding to acute asthma events for a 10 $\mu g\, m^{-3}$ increase in $PM_{2.5}$ based on data from the DAMAT study, where vague prior $\beta_{PM_{2.5}} \sim N(0, 10^2)$ was used. Panels (a) and (b) used the base prior setting as described in section 3.4. Panels (c) and (d) varied prior choices on $\boldsymbol{\nu}$ and S_t as a sensitivity analysis.

[Priors on $\boldsymbol{\nu}$ (panel (c)): A: $\nu_t = \nu^*$ for $t = 1, ..., 1096$, where $\nu^* \sim N(0, 10^2)$; B: $\nu_t \overset{iid}{\sim} N(0, 10^2)$;
G1: $\alpha \sim Gamma(0.5, 0.1)$; G2: $\alpha \sim Gamma(2, 0.2)$; G3: $\alpha \sim Gamma(10, 0.5)$; G4: $\alpha \sim Gamma(20, 1)$.
Priors on S_t (panel (d)): S1: $\pi(S'_k) \propto 1$; S2: $S'_k \overset{iid}{\sim} N(0, 10^2)$; S3: $S'_k \mid G \overset{iid}{\sim} G$, where $G \mid \alpha, G_0 \sim DP(\alpha, G_0)$.]

[TSD: time-stratified design; SBD: symmetric bidirectional design; TSLL: time-stratified log-linear; DP: Dirichlet process.]

Table 3.1: Summary of the simulation results in terms of relative bias and MSE, under different settings of temporal terms ω_t on baseline risk and exposure series X_t, with true effect size $\boldsymbol{\beta^* = 0.1}$.

	Exposure E1[a]						Exposure E2[a]					
	Baseline B1[a]		Baseline B2[a]		Baseline B3[a]		Baseline B1		Baseline B2		Baseline B3	
	RB[b](%)	MSE[b]	RB	MSE	RB	MSE	RB	MSE	RB	MSE	RB	MSE
Log-linear model (true[c])	-0.29	2.53	-0.07	2.40	-0.18	2.65	-0.43	3.15	0.37	3.43	-0.93	3.21
Log-linear model (spline[c])	-0.50	2.68	-0.11	2.59	-0.39	3.09	-0.46	3.53	0.40	3.67	-1.09	3.50
Time-stratified design												
Conditional likelihood (CLR[b])	-0.08	3.66	-0.33	3.39	-0.65	3.77	-0.44	4.29	0.92	4.33	-1.44	4.28
Bayesian (non-informative prior[d])												
Conditional likelihood	-0.10	3.67	-0.42	3.39	-0.74	3.77	-0.45	4.33	0.77	4.34	-1.49	4.32
Full likelihood with DP (T)[d]	-0.55	4.87	0.46	4.77	0.57	5.10	-1.54	6.00	-0.44	5.71	-1.25	5.60
Bayesian (informative prior[d])												
Conditional likelihood	-5.67	**2.86**[e]	-5.04	**2.71**	-4.89	2.97	-6.87	**3.27**	-5.97	3.35	-7.88	3.32
Full likelihood with DP (T)	-5.76	2.91	-5.05	2.75	-4.84	**2.87**	-7.03	3.28	-6.35	**3.15**	-7.62	**3.22**
Symmetric bi-directional design												
Conditional likelihood (CLR)	2.59	7.01	2.03	6.80	3.03	7.42	-0.75	10.87	-0.87	9.40	-2.24	9.87
Bayesian (non-informative prior)												
Conditional likelihood	2.44	7.03	1.89	6.80	2.93	7.41	-0.93	10.87	-1.05	9.39	-2.36	9.84
Full likelihood with DP (T)	2.59	7.38	2.01	7.14	2.78	7.80	-3.67	10.55	-2.93	9.26	-4.08	10.05
Bayesian (informative prior)												
Conditional likelihood	-3.64	3.72	-3.76	3.83	-3.24	3.95	-7.90	5.88	-7.91	4.39	-9.35	5.28
Full likelihood with DP (T)	-3.68	3.84	-3.93	3.72	-3.34	4.04	-8.76	5.46	-8.30	4.77	-9.16	5.31

[a] E1: auto-correlation only; E2: auto-correlation plus seasonal trend; B1: $\omega_t = \omega$; B2: $\omega_t = 0.05(1 - 0.001t)[1 + 0.5cos(2\pi t/365)]$; B3: a mixture of B2 and point mass at random spikes.

[b] RB: relative bias; MSE: mean squared error (multiplied by 10^5); CLR: conditional logistic regression.

[c] True: adjusted for true temporal trend (B1-B3); Spline: adjusted for a natural cubic spline with 7 df per year.

[d] Informative prior on β: $\beta \sim N(0.08, 0.03^2)$; Non-informative prior on β: $\beta \sim N(0, 10^2)$; DP (T): Dirichlet process prior $DP(\alpha, G_0)$ on $\boldsymbol{\nu}$ in $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$.

[e] The best performing method for a case-crossover design in terms of MSE is marked in **bold** under each baseline×exposure setting.

Table 3.2: Summary of the simulation results in terms of relative bias and MSE, under different settings of temporal terms ω_t on baseline risk and exposure series X_t, with true effect size $\boldsymbol{\beta}^* = \mathbf{1}$.

	Exposure E1[a]						Exposure E2[a]					
	Baseline B1[a]		Baseline B2[a]		Baseline B3[a]		Baseline B1		Baseline B2		Baseline B3	
	RB[b](%)	MSE[b]	RB	MSE	RB	MSE	RB	MSE	RB	MSE	RB	MSE
Log-linear model (true[c])	-0.58	0.75	-0.65	0.81	-0.80	0.96	-0.38	0.77	1.00	0.86	0.92	0.91
Log-linear model (spline[c])	-0.85	0.93	-0.90	1.06	-0.98	1.11	-0.84	1.01	1.79	1.03	1.44	1.16
Time-stratified design												
Conditional likelihood (CLR[b])	-1.92	1.01	-1.99	1.07	-2.60	1.13	-1.00	1.03	1.89	1.10	1.96	1.15
Bayesian (non-informative prior[d])												
Conditional likelihood	-1.97	1.00	-2.05	1.07	-2.63	1.14	-1.04	1.05	1.82	1.08	1.96	1.15
Full likelihood with DP[d](T)	-2.17	1.06	-1.18	1.10	-2.50	1.17	-1.25	1.03	2.35	1.15	1.57	1.18
Full likelihood with DP (N)	-2.04	1.06	-1.26	1.11	-2.43	1.17	-1.19	1.03	2.32	1.14	1.71	1.19
Full likelihood with DP (T and N)	-0.67	1.21	-0.65	1.23	-0.75	1.21	-1.00	1.09	1.46	1.29	1.10	1.36
Bayesian (informative prior[d])												
Conditional likelihood	-14.67	**0.30**[e]	-14.60	**0.31**	-15.96	0.40	-13.12	**0.36**	-14.78	**0.29**	-14.51	**0.28**
Full likelihood with DP (T)	-14.70	0.31	-14.28	**0.31**	-16.39	0.38	-13.53	0.37	-15.12	0.31	-14.80	0.30
Full likelihood with DP (N)	-14.72	0.31	-14.33	**0.31**	-16.22	**0.37**	-13.38	**0.36**	-15.13	0.31	-14.82	0.30
Full likelihood with DP (T and N)	-12.32	0.43	-11.93	0.39	-14.99	0.45	-11.63	0.43	-12.48	0.36	-12.24	0.45
Symmetric bi-directional design												
Conditional likelihood (CLR)	-5.21	1.76	-4.89	2.12	-8.34	2.43	-4.23	2.68	-3.54	2.21	-8.40	2.73
Bayesian (non-informative prior)												
Conditional likelihood	-5.71	1.78	-4.41	2.15	-8.84	2.45	-4.92	2.72	-4.09	2.25	-9.04	2.75
Full likelihood with DP (T)	-6.54	1.69	-4.79	2.04	-7.78	2.43	-6.32	2.52	-5.32	2.26	-8.25	2.83
Full likelihood with DP (N)	-6.66	1.67	-4.89	2.03	-7.67	2.41	-5.70	2.53	-5.41	2.29	-8.59	2.80
Full likelihood with DP (T and N)	-1.45	1.83	-2.43	2.10	-3.02	2.61	-1.73	2.77	-1.92	2.30	-5.27	2.84
Bayesian (informative prior)												
Conditional likelihood	-14.70	0.84	-15.60	0.87	-15.79	0.85	-17.17	0.87	-18.25	0.95	-16.12	0.96
Full likelihood with DP (T)	-14.99	0.84	-16.05	0.89	-15.42	0.92	-18.79	0.87	-18.76	0.93	-16.53	0.97
Full likelihood with DP (N)	-14.99	0.85	-16.16	0.90	-15.43	0.93	-18.77	0.87	-18.61	0.92	-16.38	0.96
Full likelihood with DP (T and N)	-12.36	0.93	-11.74	0.93	-12.90	0.99	-15.63	0.92	-17.08	0.97	-12.80	1.08

[a] E1: auto-correlation only; E2: auto-correlation plus seasonal trend; B1: $\omega_t = \omega$; B2: $\omega_t = 0.05(1 - 0.001t)[1 + 0.5cos(2\pi t/365)]$; B3: a mixture of B2 and point mass at random spikes.

[b] RB: relative bias; MSE: mean squared error (multiplied by 10^5); CLR: conditional logistic regression.

[c] True: adjusted for true temporal trend (B1-B3); Spline: adjusted for a natural cubic spline with 7 df per year.

[d] Informative prior on β: $\beta \sim N(0.8, 0.2^2)$; Non-informative prior on β: $\beta \sim N(0, 10^2)$; DP: Dirichlet process prior on the random intercepts in full likelihood formulation (T: $L^T_{full}(\beta, \boldsymbol{\nu})$; N: $L^N_{full}(\beta, \boldsymbol{\lambda})$; T and N: $L^{NT}_{full}(\beta, \boldsymbol{\lambda})$).

[e] The best performing method for a case-crossover design in terms of MSE is marked in **bold** under each baseline×exposure setting.

Table 3.3: Risk ratios of acute asthma events corresponding to a 10 $\mu g\, m^{-3}$ increase in $PM_{2.5}$ in the DAMAT study. The model was adjusted for temperature and relative humidity.

	TSD[a]			SBD[a]	
Frequentist	MLE[a]	95% CI[a]		MLE	95% CI
Conditional likelihood	1.049	(1.019, 1.080)	Conditional likelihood	1.022	(0.992, 1.052)
TSLL[a]	1.049	(1.019, 1.080)	SBLL[a]	1.022	(0.992, 1.052)
Full likelihood REM[a](T)	1.055	(1.026, 1.085)	Full likelihood REM (T)	1.020	(0.991, 1.048)
Bayesian (prior 1[b])	Bayes[a]	95% HPD[a]		Bayes	95% HPD
Conditional likelihood	1.049	(1.021, 1.081)	Conditional likelihood	1.023	(0.993, 1.053)
TSLL	1.049	(1.021, 1.081)			
Full likelihood DP[b](T)	1.055	(1.026, 1.086)	Full likelihood DP (T)	1.020	(0.992, 1.049)
Bayesian (prior 2[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.052	(1.027, 1.075)	Conditional likelihood	1.035	(1.012, 1.058)
TSLL	1.052	(1.028, 1.076)			
Full likelihood DP (T)	1.055	(1.030, 1.076)	Full likelihood DP (T)	1.034	(1.011, 1.056)
Bayesian (power prior 1[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.045	(1.024, 1.068)	Conditional likelihood	1.033	(1.010, 1.059)
TSLL	1.045	(1.024, 1.068)			
Full likelihood DP (T)	1.049	(1.025, 1.069)	Full likelihood DP (T)	1.031	(1.008, 1.058)
Bayesian (power prior 2[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.054	(1.027, 1.080)	Conditional likelihood	1.040	(1.015, 1.068)
TSLL	1.054	(1.027, 1.080)			
Full likelihood DP (T)	1.059	(1.036, 1.085)	Full likelihood DP (T)	1.041	(1.014, 1.070)
Bayesian (power prior 3[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.046	(1.018, 1.075)	Conditional likelihood	1.031	(1.007, 1.063)
TSLL	1.046	(1.017, 1.074)			
Full likelihood DP (T)	1.050	(1.022, 1.076)	Full likelihood DP (T)	1.030	(1.005, 1.063)
Bayesian (power prior 4[b])	Bayes	95% HPD		Bayes	95% HPD
Conditional likelihood	1.055	(1.027, 1.084)	Conditional likelihood	1.041	(1.012, 1.070)
TSLL	1.055	(1.027, 1.083)			
Full likelihood DP (T)	1.060	(1.030, 1.087)	Full likelihood DP (T)	1.040	(1.011, 1.069)

[a] TSD: time-stratified design; SBD: symmetric bidirectional design; TSLL: time-stratified log-linear; SBLL: symmetric bidirectional log-linear; REM: random effects model; MLE: maximum likelihood estimate (penalized pseudo-likelihood for REM); CI: confidence interval; Bayes: Bayes estimates in terms of posterior mean; HPD: highest posterior density.

[b] DP (T): Dirichlet process prior $DP(\alpha, G_0)$ on $\boldsymbol{\nu}$ in $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ under the base prior setting described in section 3.4; Prior 1: non-informative prior $\beta_{PM_{2.5}} \sim N(0, 10^2)$; Prior 2: informative prior $\beta_{PM_{2.5}} \sim N(0.05, 0.02^2)$; Power prior 1: $a_0 = 0.5$; Power prior 2: $a_0 = 1.0$; Power prior 3: $a_0 \sim Beta(20, 20)$ with mean 0.50 and variance 0.08; Power prior 4: $a_0 \sim Beta(50, 1)$ with mean 0.98 and variance 0.02.

CHAPTER IV

The role of covariate heterogeneity in the meta-analysis of gene-environment interactions on quantitative traits

4.1 Introduction

Genome-wide association studies (GWAS) provide tremendous opportunities for large-scale exploration of associations between genetic variants and complex human traits. Searching genetic associations based on GWAS has been successfully identifying many susceptibility loci for a wide spectrum of phenotypes, e.g. type 2 diabetes (T2D) (Scott et al. (2007), Zeggini et al. (2008), Morris et al. (2012), Saxena et al. (2013)), cardiovascular outcomes (Psaty et al. (2009), Sarwar et al. (2012)) and breast cancer (Song et al. 2013). Many studies have found that the risk of most complex traits are influenced by both genetic and environmental factors. The definition of 'environment' can be quite broad, including demographic factors (age, gender etc.), behavioral factors (smoking, alcohol consumption, diet, medication use etc.), and external factors (exposure to air pollution, radio-active substances etc.). The agnostic discovery strategy of GWAS may also be used to detect gene-environment interactions (GEI) that explain components of the unexplained heritability. Detecting reproducible GEI can help to further characterize the genetic architecture of human traits through sub-group or joint effects (Khoury and Wacholder (2009); Mukherjee et al. (2012)). Therefore, researchers are now looking beyond the marginal genetic effects and searching for GEI, with limited number of findings so far. For example, body

mass index (BMI) and PPARG variants appear to have a synergistic effect on fasting insulin levels and T2D (Manning et al. 2011); physical activity has been shown to attenuate the effect of fat mass associated (*FTO*) gene variants on obesity risk (Kilpeläinen et al. 2011). The associated common variants or GEIs detected so far, typically have only small to modest effects, warranting the need for large sample sizes and collaboration across different study sites for joint or meta-analysis. A number of new loci have been discovered, as well as existing loci that were initially ambiguous from individual GWAS have been validated, with high confidence through GWA meta-analysis (GWAMA) (e.g. Zeggini et al. (2008), Voight et al. (2010), Morris et al. (2012)). Many consortia have been formed to share individual level data from multiple GWAS of related traits, e.g. the DIAGRAM (T2D) (Zeggini et al. (2008), Voight et al. (2010), Morris et al. (2012)), MAGIC (glucose and insulin) (Dupuis et al. (2010), Scott et al. (2012)), CHARGE (heart and aging research) (Psaty et al. 2009), GIANT (anthropometrics) (Speliotes et al. 2010), and Global Lipids (Teslovich et al. 2010) GWAS consortia. There are also computationally efficient tools to implement GWAMA (e.g. METAL (Willer et al. 2010)). While the work on the meta-analysis of marginal genetic association effects on a binary or quantitative trait is vast, currently there are relatively few papers that explore analytical issues associated with meta-analysis of GEI (e.g. Manning et al. (2011), Aschard et al. (2011)).

There is a large variety of literature on meta-analytical technique for randomized clinical trials that can be implemented in genetic epidemiology, e.g., the fixed-effects model (FEM) (Whitehead and Whitehead 1991) and random-effects model (REM) (DerSimonian and Laird 1986). The term FEM in the classical literature (Fleiss 1993) most often refer to a model with fixed and *common* or *identical* effect. In general, a fixed-effects model only requires that there are fixed and unrelated effects in each study, irrespective of similarity of these effects across studies. However, one has to be cautious about the sci-

entific interpretation of a standard inverse-variance weighted (IVW) estimator under the general fixed-effects model where the corresponding parameters could vary across studies as opposed to a single common parameters. The REM on the other hand assumes that the parameters in each study represent random draws from a single mixing distribution. The parameters of this mixing distribution are fit by available data. The choice of FEM versus REM depends on whether the population of studies is limited to only the ones presented in meta-analysis or a larger population of studies from which the current studies are a randomly drawn sample. Even without the *common* effect assumption, assessment of effect homogeneity may be desirable in FEM. The most commonly used test of homogeneity is the Cochran's Q-test (Cochran 1954). In all our subsequent discussions, we will assume a *common* fixed effect model, consider estimation of the corresponding *common* parameter and testing the null hypotheses that this *common* parameter is zero.

The joint analysis of individual patient data (IPD) from all studies is typically regarded as the 'gold standard' for evidence synthesis. However, considerable time and resources are required to share individual level data even in an existing consortium. We refer to the joint analysis of IPD as IPD analysis (also called mega-analysis in some papers, e.g. Lin and Zeng (2010), LZ from now on), and classify the methods that combine summary statistics derived from analysis of different cohorts as meta-analysis. A natural question to ask is how much efficiency gain can be achieved by analyzing IPD over meta-analysis. Recently, LZ had considered a general multivariate inverse-variance weighted (MIVW) estimator. MIVW estimator was shown to be asymptotically equivalent to the IPD estimator, given that all the common parameters with full covariance matrix under the underlying FEM are pooled across studies in the MIVW. However, in meta-analysis of published results, it is often difficult to obtain the full covariance matrix, while univariate summary statistics (e.g. estimate and standard error) for the effects of interest are in general readily

available. LZ also characterized and quantified the efficiency loss of using an univariate IVW (UIVW) versus a MIVW estimator. The results of LZ are in a general setting but do not target towards interactions. In this chapter, we specifically focus on the estimation and testing of GEI parameter under a meta-analytic setting, and propose an adaptively weighted estimator (AWE) that uses univariate summary statistics and achieves the same asymptotic efficiency as IPD or MIVW estimator.

Another pragmatic question to ask is whether we can detect GEI from summary statistics obtained from previously conducted genome-wide meta-analysis of marginal genetic effects, without the knowledge of IPD. Meta-regression (MR) is a regression-based technique to investigate whether some particular study-level covariates explain heterogeneity among effect estimates from multiple studies. Many studies (e.g., Simmonds and Higgins (2007), SH from now on; Kovalchik (2013)) have compared aggregate data analysis (e.g. MR) with IPD analysis to detect treatment-covariate interactions for randomized clinical trials (analogous to gene-environment interactions in our case). SH showed that, under the three methods of IPD, UIVW and MR, analytical power formula to detect interactions can be expressed in terms of total, within and between study sum of squares corresponding to the covariate under certain natural assumptions. In absence of IPD, SH recommended using UIVW versus MR if the within study heterogeneity exceeds between study heterogeneity in covariates and vice versa. We borrow from their work to derive similar analytical expressions for testing GEI.

The novel adaptively weighted estimator (AWE), instead of a discrete choice of UIVW versus MR, combines UIVW and MR to archive the same asymptotic efficiency as the IPD estimator under certain conditions. The AWE has some advantages over the MIVW estimator: (1) AWE requires only univariate summary statistics from each study (study-specific estimate and standard error for the marginal association of G and GEI parameter,

and study-level mean of E); and (2) when the effect of G or E are uncommon across studies, or when covariate E is centered (very common in interaction models), MIVW will lose precision but AWE is robust to these situations.

The rest of the chapter is organized as follows. In the methods section we describe different strategies for meta-analysis of GEI, followed by analytical results on bias, variance and power properties of these estimators. A comprehensive simulation study was performed to assess the performance of the meta-analysis methods under a variety of scenarios. We primarily explore the issue of covariate heterogeneity, but also explore several other important factors that could potentially affect the relative performance of these methods: (1) departures from gene-environment (GE) independence; (2) heterogeneity in minor allele frequencies (MAFs) across cohorts; (3) lack of a common set of confounders to adjust for in individual studies; (4) misspecification of the genetic susceptibility model (dominant/co-dominant/additive); and finally (5) the presence of a non-linear form of interaction. In the results section, we report simulation findings followed by an illustrative example, where we examine whether Single Nucleotide Polymorphisms (SNPs) in *FTO* gene modify the effect of environmental factors (age and BMI) on high-density lipoprotein cholesterol (HDL-C) levels, a T2D related quantitative trait. We hope this chapter to provide useful insight and guidelines while conducting meta-analysis of GEI.

4.2 Methods

4.2.1 Detecting GEI via meta-analysis

Consider a quantitative trait Y, a continuous environmental exposure E and a bi-allelic genetic locus G with genotypes of AA, Aa and aa, where A is the minor allele. Suppose that there are K independent studies and a total of N participants, with n_k participants in the k-th study, $k = 1, ..., K$, $\sum_{k=1}^{K} n_k = N$. Let Y_{ki}, E_{ki} and G_{ki} be the pheno-

type, environmental exposure and genotype for participant i in study k respectively, for $i = 1, ..., n_k, k = 1, ..., K$. We consider the possibility of adjusting for demographic covariates/confounders $\boldsymbol{Z}$ in the model. The true model for individual responses follows the FEM

$$(4.1)\quad Y_{ki} = \alpha_k + \beta_G G_{ki} + \beta_E E_{ki} + \delta G_{ki} E_{ki} + \boldsymbol{\beta}_Z^\top \boldsymbol{Z}_{ki} + \epsilon_{ki}, \quad i = 1, ..., n_k,\ k = 1, ..., K,$$

where α_k is the study specific intercept, β_G and β_E are the main effects corresponding to the genetic factor and environmental exposure respectively, δ is the GEI effect of interest, and $\boldsymbol{\beta}_Z$ is the effect of covariates $\boldsymbol{Z}$. $\boldsymbol{\beta} = (\beta_G, \beta_E, \delta, \boldsymbol{\beta}_Z)$ is assumed to be fixed and common across studies under model (4.1). We will further discuss certain situations where the 'common' effect assumption can be relaxed in later sections. The random errors ϵ_{ki}'s were assumed as $\epsilon_{ki} \sim N(0, \sigma_k^2)$. Various susceptibility models including dominant model ($G = 1$ if AA and Aa; $G = 0$ if aa), recessive model ($G = 1$ if AA; $G = 0$ if Aa and aa), additive model ($G = 2$ if AA; $G = 1$ if Aa; $G = 0$ if aa) and co-dominant model ($G = AA, Aa$ or aa with aa as the reference level) are considered. For co-dominant models, $\boldsymbol{\beta}_G = (\beta_G^{Aa}, \beta_G^{AA})$ and $\boldsymbol{\delta} = (\delta^{Aa}, \delta^{AA})$ in model (4.1), corresponding to genotypes Aa and AA respectively.

We first describe three traditional approaches to detect GEI under model (4.1) but without $\boldsymbol{Z}$ to simplify the presentation, including IPD analysis, standard meta-analysis (using UIVW or MIVW), and MR. For the sake of completeness, we also describe REM meta-analysis as well as a two-step estimator previously suggested by SH. We then propose an AWE that combines UIVW and MR estimator, which is shown to be an unbiased and fully efficient estimator as the IPD estimator under certain plausible assumptions, but using only univariate summary statistics.

Existing methods

(i) Individual patient data analysis: The IPD analysis fits model (4.1) using the individual level data. The weighted least square (WLS) method can be used to deal with the heterogeneous σ_k^2 across studies. Instead, for our IPD analysis, we assume $\sigma_k^2 = \sigma^2$ for $k = 1, ..., K$. Let $\boldsymbol{X}$ be the design matrix and $\boldsymbol{Y}$ be the response in model (4.1). The maximum likelihood estimate (MLE) of δ under linear regression model (4.1) (denote as $\hat{\delta}^{\text{IPD}}$) can be obtained through the corresponding element of $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}$, and its estimated variance $\hat{\mathbf{v}}(\hat{\delta}^{\text{IPD}})$ can be obtained through the corresponding diagonal element of $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\hat{\sigma}^2$. Throughout this chapter, we use the generic notation $\mathbf{v}(\hat{\delta})$ for the asymptotic model based variance (covariance matrix for multivariate $\hat{\boldsymbol{\delta}}$) of any given estimator $\hat{\delta}$, and $\hat{\mathbf{v}}(\hat{\delta})$ for the corresponding estimated variance. We will present some simplified expressions for $\hat{\mathbf{v}}(\hat{\delta})$ under specific structures of $\boldsymbol{X}^\top \boldsymbol{X}$ in section 4.2.2.

(ii) Meta-analysis using inverse-variance weighted estimator: Since the data required to perform IPD analysis are often not available in published results, meta-analysis that combines summary statistics from individual studies may be what is practically feasible. We consider some variants of IVW estimator under the FEM (4.1).

(ii.A) UIVW: A UIVW estimator needs the collection of the MLEs $\hat{\delta}_k$ and $\hat{\mathbf{v}}(\hat{\delta}_k)$ estimated from model (4.1) using data from only study k. A FEM assumes that $\hat{\delta}_k \overset{iid}{\sim} N(\delta, \mathbf{v}(\hat{\delta}_k))$, where $\mathbf{v}(\hat{\delta}_k)$ is the model based asymptotic variance of $\hat{\delta}_k$. Let $\boldsymbol{X}_k$ be the design matrix and $\boldsymbol{Y}_k$ be the response of study k. $\hat{\delta}_k$ and $\hat{\mathbf{v}}(\hat{\delta}_k)$ can be obtained through the corresponding element of $(\boldsymbol{X}_k^\top \boldsymbol{X}_k)^{-1}\boldsymbol{X}_k^\top \boldsymbol{Y}_k$ and $(\boldsymbol{X}_k^\top \boldsymbol{X}_k)^{-1}\hat{\sigma}_k^2$ respectively. Then the UIVW estimator under a FEM is given by

$$\hat{\delta}^{\text{UIVW}} = \Big\{\sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1}\Big\}^{-1} \sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1}\hat{\delta}_k \text{ with variance } \hat{\mathbf{v}}(\hat{\delta}^{\text{UIVW}}) = \Big\{\sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1}\Big\}^{-1}.$$

The validity of the method requires certain 'standard condition', namely: for a large study

k, $\hat{\delta}_k$ is asymptotically normal $\hat{\delta}_k \overset{iid}{\sim} N(\delta, \mathbf{v}(\hat{\delta}_k))$ and the true asymptotic variance $\mathbf{v}(\hat{\delta}_k)$ can be estimated by $\hat{\mathbf{v}}(\hat{\delta}_k)$ with negligible error (Whitehead and Whitehead 1991). We refer to the condition as 'standard' throughout, and we note that it is often implicitly assumed to hold in classic meta-analysis literature (e.g., DerSimonian and Laird (1986), Whitehead and Whitehead (1991), LZ).

(ii.B) MIVW: Let $\hat{\boldsymbol{\beta}}_k = (\hat{\beta}_{Gk}, \hat{\beta}_{Ek}, \hat{\delta}_k)$ be the MLE of $\boldsymbol{\beta}$ from study k with estimated variance $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)$ obtained through the corresponding sub-matrix of $(\boldsymbol{X}_k^\top \boldsymbol{X}_k)^{-1}\hat{\sigma}_k^2$. When both $\hat{\boldsymbol{\beta}}_k$ and $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)$ are available from each study, we consider the MIVW estimator following LZ,

$$\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}} = \Big\{\sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1}\Big\}^{-1} \sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1}\hat{\boldsymbol{\beta}}_k \text{ with variance } \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}}) = \Big\{\sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1}\Big\}^{-1}.$$

Then $\hat{\delta}^{\mathrm{MIVW}}$ and $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MIVW}})$ corresponding to the interaction parameter δ can be obtained from the corresponding element of $\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}}$ and $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}})$ respectively. LZ showed that $\hat{\delta}^{\mathrm{MIVW}}$ has full asymptotic efficiency as $\hat{\delta}^{\mathrm{IPD}}$ under a FEM. However, $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)$ may be difficult to acquire in meta-analysis of published results, and $\hat{\delta}^{\mathrm{UIVW}}$ is the one most commonly used.

(ii.C) REM: Alternatively, if the set of studies are considered to be a random sample from the population of studies, one can use a REM assuming $\hat{\delta}_k|\delta_k \overset{iid}{\sim} N(\delta_k, \mathbf{v}(\hat{\delta}_k))$, where $\delta_k|\delta, \tau^2 \sim N(\delta, \tau^2)$ and $\mathbf{v}(\hat{\delta}_k)$ is the same asymptotic variance used in UIVW. Following DerSimonian and Laird (1986),

$$\begin{aligned}
\hat{\delta}^{\mathrm{REM}} &= \Big[\sum_k \big\{\hat{\tau}^2 + \hat{\mathbf{v}}(\hat{\delta}_k)\big\}^{-1}\Big]^{-1}\Big[\sum_k \big\{\hat{\tau}^2 + \hat{\mathbf{v}}(\hat{\delta}_k)\big\}^{-1}\hat{\delta}_k\Big], \\
\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{REM}}) &= \Big[\sum_k \big\{\hat{\tau}^2 + \hat{\mathbf{v}}(\hat{\delta}_k)\big\}^{-1}\Big]^{-1},
\end{aligned}$$

where $\hat{\tau}^2 = \max\Big[0, \frac{Q-(K-1)}{\sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1} - \{\sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-2}/\sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1}\}}\Big]$. Here $Q = \sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1}(\hat{\delta}^{\mathrm{UIVW}} - \hat{\delta}_k)^2$ is the Cochran's Q statistic (Cochran (1954)) used to test homogeneity of δ_k, i.e., $H_0 : \delta_1 = \delta_2 = \cdots = \delta_K$. Note that, if $\hat{\tau}^2 = 0$, $\hat{\delta}^{\mathrm{REM}} = \hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{REM}}) = \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})$.

(iii) Meta-regression: The IPD model (4.1) implies that the marginal genetic effect depends linearly on E. We consider a linear MR model to reveal the underlying dependence between the marginal genetic effects and the study mean values of E (say $m_k = \sum_i E_{ki}/n_k$). Screening for the marginal effect of G is routinely performed as the first step in GWA analysis. For each study k, we first consider the marginal genetic association model

$$Y_{ki} = \lambda_{0k} + \lambda_k G_{ki} + \eta_{ki}, \quad i = 1, ..., n_k. \tag{4.2}$$

where the errors $\eta_{ki} \sim N(0, \sigma^2_{\eta k})$. At the second step, the MLE $\hat{\lambda}_k$ is regressed on m_k through the MR model

$$\hat{\lambda}_k = \gamma_0 + \gamma m_k + \nu_k, \quad k = 1, ..., K. \tag{4.3}$$

Denote $\hat{\mathbf{v}}(\hat{\lambda}_k)$ as the model based variance estimated from (4.2). To account for the potential heterogeneity in $\hat{\mathbf{v}}(\hat{\lambda}_k)$ across studies, we consider the WLS estimator of γ (denoted as $\hat{\delta}^{\mathrm{MR}}$) in model (4.3) with weight $w_k = \hat{\mathbf{v}}(\hat{\lambda}_k)^{-1}$ assumed as known, i.e., $\nu_k \sim N(0, \hat{\mathbf{v}}(\hat{\lambda}_k))$. Let $\overline{m} = (\sum_k w_k m_k)/(\sum_k w_k)$, $\hat{\delta}^{\mathrm{MR}}$ and $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})$ can be derived as

$$\begin{aligned} \hat{\delta}^{\mathrm{MR}} &= \{\sum_k w_k (m_k - \overline{m})^2\}^{-1} \{\sum_k w_k (m_k - \overline{m}) \hat{\lambda}_k\}, \\ \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}}) &= \{\sum_k w_k (m_k - \overline{m})^2\}^{-1}. \end{aligned}$$

The advantage of MR approach is that one can identify GEI with only limited summary data on E (only the mean m_k's) and published results of marginal genetic effects ($\hat{\lambda}_k$ and $\hat{\mathbf{v}}(\hat{\lambda}_k)$).

(iv) Two-stage estimator: Let $m = \sum_{k,i} E_{ki}/N$ denote the overall sample mean of E, $s^2_E = N^{-1} \sum_{k,i} (E_{ki} - m)^2$ denote the total sample variance of E, and $s^2_{Ek} = n_k^{-1} \sum_{i=1}^{n_k} (E_{ki} - m_k)^2$ denote the sample variance of E within the k-th study. Denote the population parameters for m, m_k, s^2_E, s^2_{Ek} as μ, μ_k, σ^2_E, σ^2_{Ek} respectively. We make the usual par-

tition of the total sum of squares (TSS) of E as the sum of the within-study sum of squares (WSS) and between-study sum of squares (BSS), i.e., $TSS = WSS + BSS$, where $TSS = \sum_{k,i}(E_{ki} - m)^2 = Ns_E^2$, $WSS = \sum_k \sum_i (E_{ki} - m_k)^2 = \sum_k n_k s_{Ek}^2$ and $BSS = \sum_k n_k (m_k - m)^2$. Throughout this chapter, we assume $n_k/N \to r_k \in (0, 1)$ as $N \to \infty$. Consider the estimands $tss = \sigma_E^2$, $wss = \sum_k r_k \sigma_{Ek}^2$ and $bss = \sum_k r_k (\mu_k - \mu)^2$. We have $TSS/N \xrightarrow{p} tss$, $WSS/N \xrightarrow{p} wss$, $BSS/N \xrightarrow{p} bss$, as $N \to \infty$.

Inspired by our analytical result (shown later in section 4.2.2) that the asymptotic relative efficiency (ARE) between $\hat{\delta}^{\mathrm{MR}}$ and $\hat{\delta}^{\mathrm{UIVW}}$ is bss/wss, we define $Q_E = BSS/WSS$ as a statistic for measuring heterogeneity of E between studies relative to the within study heterogeneity, and consider a two-stage approach

$$\hat{\delta}^{\mathrm{TS}} = \begin{cases} \hat{\delta}^{\mathrm{UIVW}}, & \text{if } Q_E \leq 1; \\ \hat{\delta}^{\mathrm{MR}}, & \text{if } Q_E > 1, \end{cases}$$

i.e., using $\hat{\delta}^{\mathrm{UIVW}}$ instead of $\hat{\delta}^{\mathrm{MR}}$ if $WSS \geq BSS$ and vice versa. This is an ad-hoc procedure of discretely determining which method to use. Q_E is one of the two-stage test statistics suggested in SH.

Adaptively weighted estimator

We note that, using only summary statistics, both $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$ can potentially lack precision. Moreover, $\hat{\delta}^{\mathrm{MR}}$ can be subject to significant ecological bias (Morgenstern (1982), Greenland (1987), Schwartz (1994), Berlin et al. (2002)). Thus, we propose an adaptive estimator that combines $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$ to improve efficiency. We first prove the following lemma.

Lemma 1. *Let Y_i be independent random variables with equal variance, for $i = 1, ..., n$, and let $X_j = (X_{1j}, ..., X_{nj})^\top$ be the j-th predictor, $j = 1, ..., p + q$. Let $\hat{\zeta}_j$ ($j = 1, ..., p$) and $\hat{\theta}_j$ ($j = 1, ..., p + q$) be the MLEs of the parameters under the two nested linear*

regression models

$$Y_i = \zeta_0 + \sum_{j=1}^{p} \zeta_j X_{ij} + \eta_i \quad \textit{and} \quad Y_i = \theta_0 + \sum_{j=1}^{p+q} \theta_j X_{ij} + \epsilon_i,$$

then $\hat{\boldsymbol{\zeta}} = (\hat{\zeta}_1, ..., \hat{\zeta}_p)$ *and* $\hat{\boldsymbol{\theta}}_2 = (\hat{\theta}_{p+1}, ..., \hat{\theta}_{p+q})$ *are asymptotically independent.*

Proof of Lemma 1 is presented in Appendix B.3. The notations used in Lemma 1 are generic and unrelated to the ones defined elsewhere in this chapter. Applying Lemma 1 to model (4.1) and (4.2), the marginal genetic association $\hat{\lambda}_k$ and GEI $\hat{\delta}_k$ are asymptotically independent for each study k, as they are coming from two nested linear regression models. Note that $\hat{\delta}^{\mathrm{UIVW}}$ is a linear combination of $\hat{\delta}_k$, and that $\hat{\delta}^{\mathrm{MR}}$ is a linear combination of $\hat{\lambda}_k$, then the following corollary holds.

Corollary 4.1. $\hat{\delta}^{\mathrm{UIVW}}$ *and* $\hat{\delta}^{\mathrm{MR}}$ *are asymptotically independent.*

Borrowing the classic idea of an IVW estimator along with the standard condition, we propose an AWE of the form

$$\hat{\delta}^{\mathrm{AWE}} = \{\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})^{-1}\}^{-1}\{\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})^{-1}\hat{\delta}^{\mathrm{UIVW}} + \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})^{-1}\hat{\delta}^{\mathrm{MR}}\},$$

which combines $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$ using the inverse-variances as weights. In order to calculate $\hat{\delta}^{\mathrm{AWE}}$, summary statistics of study-specific effect estimates ($\hat{\delta}_k$, $\hat{\mathbf{v}}(\hat{\delta}_k)$, $\hat{\lambda}_k$ and $\hat{\mathbf{v}}(\hat{\lambda}_k)$) and study-level covariate means m_k are needed from each study k. The intuitive rationale behind the AWE is that, when $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})$ is relatively smaller than $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})$, $\hat{\delta}^{\mathrm{AWE}}$ puts more weight on $\hat{\delta}^{\mathrm{UIVW}}$ and vice versa. The estimated weights can be translated in terms of the ratio of WSS versus BSS as presented in section 4.2.2. Theorem 4.1 establishes that with this particular choice of weights, $\hat{\delta}^{\mathrm{AWE}}$ has the maximal precision within the class of weighted estimators of the form $\hat{\delta}^{\mathrm{AWE}}(w) = w\hat{\delta}^{\mathrm{UIVW}} + (1-w)\hat{\delta}^{\mathrm{MR}}$, $0 \leq w \leq 1$.

Theorem 4.1. *For the class of weighted estimators* $\hat{\delta}^{\mathrm{AWE}}(w) = w\hat{\delta}^{\mathrm{UIVW}} + (1-w)\hat{\delta}^{\mathrm{MR}}$, $0 \leq w \leq 1$, $\mathbf{v}(\hat{\delta}^{\mathrm{AWE}}(w))^{-1}$ *attains its maximum at* $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}$ *if and only if the weight* $w = \mathbf{v}(\hat{\delta}^{\mathrm{MR}})/\{\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})\}$.

Proof of Theorem 4.1 is presented in Appendix B.3. A consequence of Theorem 4.1 is that the precision of $\hat{\delta}^{\mathrm{AWE}}$ is the sum of the precisions of $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$. Under the standard condition, $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{AWE}})^{-1} = \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})^{-1}$. We will further show that $\hat{\delta}^{\mathrm{AWE}}$ is fully efficient as $\hat{\delta}^{\mathrm{IPD}}$ under certain natural assumptions described in section 4.2.2.

Remark 4.1: Co-dominant model. For the co-dominant model that $\boldsymbol{\delta} = (\delta^{Aa}, \delta^{AA})$, it is straightforward to modify the proposed methods to their bivariate counterparts. In particular, $\hat{\boldsymbol{\delta}}^{\mathrm{IPD}}$ and $\hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}^{\mathrm{IPD}})$ can be obtained from (4.1); $\hat{\boldsymbol{\delta}}^{\mathrm{UIVW}}$ and $\hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}^{\mathrm{UIVW}})$ can be obtained as $\left\{\sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}_k)^{-1}\right\}^{-1} \sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}_k)^{-1}\hat{\boldsymbol{\delta}}_k$ and $\left\{\sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}_k)^{-1}\right\}^{-1}$; $\hat{\boldsymbol{\delta}}^{\mathrm{MIVW}}$ and $\hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}^{\mathrm{MIVW}})$ can be obtained from $\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}}$ and $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}})$; MR model can be modified as a multiple response regression $\hat{\boldsymbol{\lambda}}_k = \boldsymbol{\gamma}_0 + \boldsymbol{\gamma} m_k + \boldsymbol{\nu}_k$, where $\hat{\boldsymbol{\lambda}}_k = (\hat{\lambda}_k^{Aa}, \hat{\lambda}_k^{AA})^\top$ and $\nu_k \overset{iid}{\sim} N(0, \hat{\mathbf{v}}(\hat{\boldsymbol{\lambda}}_k))$. Corollary 4.1 and Theorem 4.1 also hold following Lemma 1 for bivariate $\boldsymbol{\delta}$. A bivariate form of AWE can be considered as $\hat{\boldsymbol{\delta}}^{\mathrm{AWE}} = \{\hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}^{\mathrm{UIVW}})^{-1} + \hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}^{\mathrm{MR}})^{-1}\}^{-1}$ $\{\hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}^{\mathrm{UIVW}})^{-1}\hat{\boldsymbol{\delta}}^{\mathrm{UIVW}} + \hat{\mathbf{v}}(\hat{\boldsymbol{\delta}}^{\mathrm{MR}})^{-1}\hat{\boldsymbol{\delta}}^{\mathrm{MR}}\}$.

4.2.2 Analytical results

This section shows the analytical results regarding bias, variance and power to provide theoretical support for the proposed methods described in section 4.2.1.

Bias

Following classic linear regression and meta-analysis results, $\hat{\delta}^{\mathrm{IPD}}$, $\hat{\delta}^{\mathrm{UIVW}}$, $\hat{\delta}^{\mathrm{REM}}$ and $\hat{\delta}^{\mathrm{MIVW}}$ are all asymptotically unbiased estimators of δ. However, $\hat{\delta}^{\mathrm{MR}}$ is not necessarily unbiased for δ in general. The relationship between the marginal effect of G and the study-specific means m_k may differ from the underlying relationship between the marginal

effect of G and individual level data for E. This phenomenon was termed as 'ecological bias' or 'ecological fallacy', and well characterized in the literature (Morgenstern (1982), Greenland (1987), Schwartz (1994)). However, we note that $\hat{\delta}^{\mathrm{MR}}$ is an unbiased estimator of δ under the following GE independence assumption. We use the generic notation $P(\cdot)$ to denote the distribution of a random variable.

Assumption 4.1. *$P(G, E|\, study = k) = P(G|\, study = k)P(E|\, study = k)$, for $k = 1, ..., K$, i.e., G and E are independent within each study.*

Proposition 4.1. *Under assumption 4.1, $\hat{\delta}^{\mathrm{MR}}$ of model (4.3) is asymptotically unbiased for δ.*

Proof of Proposition 4.1 is presented in Appendix B.3. Proposition 4.1 holds for dominant, recessive, additive and co-dominant genetic susceptibility models. If the independence assumption is relaxed, a slightly more complex MR model can lead to an unbiased estimator of δ under each susceptibility model (shown in Appendix B.3). Our simulation study shows that the bias of $\hat{\delta}^{\mathrm{MR}}$ is not to a level of practical concern under the current setting of a linear interaction between G and continuous E, even when there is evidence of departure from GE independence. Therefore, we focus on model (4.3) in order to demonstrate our results under a global and simple MR model for all the susceptibility models regardless of GE dependence.

Remark 4.2: Bias of $\hat{\delta}^{\mathrm{MR}}$ in terms of tss/bss. Without assumption 4.1, we showed (under certain assumptions in Appendix B.3) that the limiting value of the bias of $\hat{\delta}^{\mathrm{MR}}$ is proportional to the ratio tss/bss and the correlation between G and E. If the correlations within each study is 0, then $E(\hat{\delta}^{\mathrm{MR}}) - \delta \xrightarrow{p} 0$. If assumption 4.1 holds, $\hat{\delta}^{\mathrm{AWE}}$ is an asymptotically unbiased estimator of δ under the standard condition, as both components are unbiased. Moreover, we showed later that the limiting value of the weight corresponding

to $\hat{\delta}^{\mathrm{MR}}$ in $\hat{\delta}^{\mathrm{AWE}}$ is bss/tss. So $\hat{\delta}^{\mathrm{AWE}}$ adaptively puts less weight on $\hat{\delta}^{\mathrm{MR}}$ when the bias of $\hat{\delta}^{\mathrm{MR}}$ increases.

Variance

Explicit variance formulae $\hat{\mathbf{v}}(\hat{\delta})$ as well as the corresponding asymptotic variance $\mathbf{v}(\hat{\delta})$ for each estimator was derived under GE independence assumption (shown in Appendix B.3). Because the simple linear regression likelihood $\prod_{k,i} P(Y_{ki}|G_{ki}, E_{ki})$ corresponding to model (4.1) does not use any assumptions about the joint stochastic distribution of G and E, the role of the GE independence assumption in this chapter is only to provide the explicit variance expression. This is different from case-control studies where assuming GE independence and using the retrospective likelihood leads to huge gain in efficiency (Piegorsch et al. (1994), Umbach and Weinberg (1997), Chatterjee and Carroll (2005)).

In this section, we assume $\sigma_k^2 = \sigma^2$ for $k = 1, ..., K$, and consider a dominant susceptibility model for stating Theorems 4.2 and 4.3. We discuss extension to additive and co-dominant models later in this section. Let $G = 1$ ($G = 0$) indicate whether an individual is a carrier (non-carrier) of the minor allele A, and let p_k denote $P(G = 1|\,\text{study} = k)$ the carrier frequencies in study k, $k = 1, \cdots, K$.

Theorem 4.2. *Under assumption 4.1,*

$\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} \geq \mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{AWE}})^{-1}$. *The equality holds if and only if* $p_k = p$, *for* $k = 1, 2, ..., K$, *where* p *is the common carrier frequencies across all studies.*

Proof of Theorem 4.2 is shown in Appendix B.3. Under assumption 4.1, the precision of $\hat{\delta}^{\mathrm{IPD}}$ is in general greater than that of $\hat{\delta}^{\mathrm{AWE}}$. However, under the additional assumption of homogeneity of the MAFs, we have $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}}) = \mathbf{v}(\hat{\delta}^{\mathrm{AWE}})$. We call this assumption 4.2.

Assumption 4.2. *The MAFs corresponding to the susceptible SNP are constant across all studies, i.e.* $p_k = p$, *for* $k = 1, 2, ..., K$.

Theorem 4.3. *Under assumptions 4.1 and 4.2,*

$\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{AWE}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}$*, where*

$\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) = \{Np(1-p)wss\}^{-1}\sigma^2$, $\mathbf{v}(\hat{\delta}^{\mathrm{MR}}) = \{Np(1-p)bss\}^{-1}\sigma^2$ *and*

$\mathbf{v}(\hat{\delta}^{\mathrm{IPD}}) = \mathbf{v}(\hat{\delta}^{\mathrm{AWE}}) = \{Np(1-p)tss\}^{-1}\sigma^2$.

Proof of Theorem 4.3 is shown in Appendix B.3. Following Theorem 4.3, the asymptotic model based variances $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})$, $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})$, $\mathbf{v}(\hat{\delta}^{\mathrm{MR}})$ and $\mathbf{v}(\hat{\delta}^{\mathrm{AWE}})$ are all translated in terms of covariate heterogeneity of E. ARE between $\hat{\delta}^{\mathrm{UIVW}}$ ($\hat{\delta}^{\mathrm{MR}}$) and $\hat{\delta}^{\mathrm{IPD}}$ is wss/tss (bss/tss). $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) \leq \mathbf{v}(\hat{\delta}^{\mathrm{MR}})$, if $wss \geq bss$, and vice versa. For the extreme case, when there is no between-study heterogeneity in the study means of E (i.e. $\mu_k = \mu$), $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) = \mathbf{v}(\hat{\delta}^{\mathrm{IPD}})$; in contrast, if all $\sigma^2_{Ek} = 0$ (i.e. E is constant within each study), $\mathbf{v}(\hat{\delta}^{\mathrm{MR}}) = \mathbf{v}(\hat{\delta}^{\mathrm{IPD}})$. This result is consistent with LZ, as their equality condition for $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) = \mathbf{v}(\hat{\delta}^{\mathrm{IPD}})$ reduces to $\mu_k = \mu$ in our case.

The limiting weight of $\hat{\delta}^{\mathrm{AWE}}$ can be simplified as $w = \mathbf{v}(\hat{\delta}^{\mathrm{MR}})/\{\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})\} = bss^{-1}/\{wss^{-1} + bss^{-1}\} = wss/tss$. Since $WSS/TSS \xrightarrow{p} wss/tss$ and $BSS/TSS \xrightarrow{p} bss/tss$, as $N \to \infty$, we can use the estimated weights WSS/TSS and BSS/TSS in $\hat{\delta}^{\mathrm{AWE}}$, which leads to

$$\hat{\delta}^{\mathrm{AWE}} = \frac{WSS}{TSS}\,\hat{\delta}^{\mathrm{UIVW}} + \frac{BSS}{TSS}\,\hat{\delta}^{\mathrm{MR}}.$$

$\hat{\delta}^{\mathrm{AWE}}$ adaptively captured the precision trade-off between the two estimators: $\hat{\delta}^{\mathrm{AWE}}$ puts more weight on $\hat{\delta}^{\mathrm{UIVW}}$ if WSS is relatively larger than BSS, and vice versa. In summary, under assumptions 4.1 and 4.2, $\hat{\delta}^{\mathrm{AWE}}$ is consistent, unbiased, and asymptotically fully efficient estimator, which uses only univariate summary statistics without the knowledge of the original IPD. The operating characteristics for the proposed meta-analytic methods are summarized in Table 4.1. When assumption 4.1 or 4.2 is relaxed, the statements in Theorems 4.2 and 4.3 are numerically evaluated through a comprehensive simulation study.

Remark 4.3: Additive and co-dominant models. Following LZ, asymptotically $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}}) \geq \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}}) = \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MIVW}})$. One sufficient condition for the equality is $m_k = m$, i.e., when $\hat{\delta}^{\mathrm{MR}}$ undefined. In general, we have difficulties to show the analytical results of $\hat{\delta}^{\mathrm{AWE}}$ in Theorems 4.2 and 4.3 for additive and co-dominant models, though we can directly translate Theorems 4.2 and 4.3 for δ^{Aa} and δ^{AA} respectively under a co-dominant model if we use $diag(\hat{\mathbf{v}}(\hat{\lambda}_k^{Aa}), \hat{\mathbf{v}}(\hat{\lambda}_k^{AA}))$ for $\hat{\mathbf{v}}(\hat{\boldsymbol{\lambda}}_k)$ in the MR model, i.e., two separated MRs. The statements in Theorems 4.2 and 4.3 are numerically evaluated for additive and co-dominant models through comprehensive simulation studies with/without assumption 4.1 and 4.2.

Remark 4.4: Centering of covariate E. Centering is often made for continuous E to facilitate the interpretation of β_G as the main effect of G at the mean value of E. Under a meta-analysis set-up, it is natural to consider each study k has E centered at study specific mean m_k and fit the model $Y_{ki} = \alpha_k' + \beta_G' G_{ki} + \beta_E' E_{ki}' + \delta' G_{ki} E_{ki}' + \epsilon_{ki}, i = 1, ..., n_k$, where $E_{ki}' = E_{ki} - m_k$. When the IPD are available, it is natural to consider that E is centered at the overall mean m (an IPD analysis using data centered at study mean m_k for each study is not valid) and fit the model $Y_{ki} = \alpha_k^\star + \beta_G^\star G_{ki} + \beta_E^\star E_{ki}^\star + \delta^\star G_{ki} E_{ki}^\star + \epsilon_{ki}, i = 1, ..., n_k, k = 1, ..., K$, where $E_{ki}^\star = E_{ki} - m$. In this new parametrization, we have $(\beta_E', \delta') = (\beta_E, \delta)$ and $\beta_G' = \beta_G + m_k\delta$ depending on k, denoted as β_{Gk}'; $(\beta_E^\star, \delta^\star) = (\beta_E, \delta)$ and $\beta_G^\star = \beta_G + m\delta^\star$. It is clear that $\hat{\delta}^{\mathrm{IPD}}$, $\hat{\delta}^{\mathrm{UIVW}}$, $\hat{\delta}^{\mathrm{REM}}$, $\hat{\delta}^{\mathrm{MR}}$ and $\hat{\delta}^{\mathrm{AWE}}$ remain invariant with centered E since $\delta' = \delta^\star = \delta$. Thus, results analogous to Theorems 4.1-4.3 concerning $\hat{\delta}^{\mathrm{IPD}}$, $\hat{\delta}^{\mathrm{UIVW}}$, $\hat{\delta}^{\mathrm{MR}}$ and $\hat{\delta}^{\mathrm{AWE}}$ also hold for the centered models. However, results corresponding to $\hat{\delta}^{\mathrm{MIVW}}$ need to be modified. Denote $\hat{\delta}^{\mathrm{MIVW}'}$ as the MIVW estimator obtained by pooling $(\beta_{Gk}', \beta_E', \delta')$ with 3×3 covariance matrix and denote $\hat{\delta}^{\mathrm{MIVW2}'}$ as the MIVW estimator obtained by pooling the two common effects (β_E', δ') with 2×2 covariance matrix from the centered model. We can show that $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) \geq \mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'}) = \mathbf{v}(\hat{\delta}^{\mathrm{MIVW}'}) \geq \mathbf{v}(\hat{\delta}^{\mathrm{IPD}})$ (in Appendix B.3). These

results are consistent with LZ, as the true model has three common fixed-effects and there is efficiency loss by pooling a subset of the common parameters. For the centered model, the MIVW estimator is not fully efficient. A solution is to consider an alternative AWE as $\hat{\delta}^{\mathrm{AWE2'}} = \{\mathbf{v}(\hat{\delta}^{\mathrm{MR}})\hat{\delta}^{\mathrm{MIVW2'}} + \mathbf{v}(\hat{\delta}^{\mathrm{MIVW2'}})\hat{\delta}^{\mathrm{MR}}\}/\{\mathbf{v}(\hat{\delta}^{\mathrm{MIVW2'}}) + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})\}$. Under assumption 4.1, we have (shown in Appendix B.3) $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{AWE2'}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{MIVW2'}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1} \geq \mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{AWE}})^{-1}$. The equality holds if and only if $p_k = p$, for $k = 1, 2, ..., K$. So $\hat{\delta}^{\mathrm{AWE2'}}$ is fully efficient under assumption 4.1.

Remark 4.5: Fixed and uncommon parameters (β_{Gk}, β_{Ek}). When the uncommon parameters (β_{Gk}, β_{Ek}) and a common δ are considered in the true IPD model (4.1), it is clear that results regarding UIVW and MR still hold as they focus only on the common parameter δ (MR model (4.3) needs to be changed as $\hat{\lambda}_k = \gamma_{0k} + \gamma m_k + \nu_k$). However, according to discussion in Remark 4.4, $\hat{\delta}^{\mathrm{MIVW}}$ would lose precision since (β_{Gk}, β_{Ek}) are uncommon across studies. Actually, $\mathbf{v}(\hat{\delta}^{\mathrm{MIVW}}) = \mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})$ in this case. Moreover, we have $\mathbf{v}(\hat{\delta}^{\mathrm{AWE}}) < \mathbf{v}(\hat{\delta}^{\mathrm{MIVW}})$ since $\hat{\delta}^{\mathrm{AWE}}$ has the sum of the precision of $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$, when the effect of G or E are uncommon across studies.

Power

For dominant and additive models, we consider the Wald-type test statistic $T = \hat{\mathbf{v}}(\hat{\delta})^{-\frac{1}{2}}\hat{\delta}$ for testing the null hypothesis H_0: $\delta = 0$ against H_1: $\delta \neq 0$. The power to detect an effect size δ^* at level α is approximately $P_w(\hat{\delta}, \delta^*) = \Phi(-z_{\frac{\alpha}{2}} + \hat{\mathbf{v}}(\hat{\delta})^{-\frac{1}{2}}\delta^*) + \Phi(-z_{\frac{\alpha}{2}} - \hat{\mathbf{v}}(\hat{\delta})^{-\frac{1}{2}}\delta^*)$, where Φ is the cumulative distribution function (CDF) of a standard normal variable z and $z_{\frac{\alpha}{2}}$ is the corresponding $\frac{\alpha}{2}$ th upper percentile. For co-dominant models, we consider a joint Wald test statistic $T = \hat{\boldsymbol{\delta}}^{\top}\hat{\mathbf{v}}(\hat{\boldsymbol{\delta}})^{-1}\hat{\boldsymbol{\delta}} \overset{H_0}{\sim} \chi^2_2$ for testing H_0: $\boldsymbol{\delta} = \mathbf{0}$ against H_1: $\boldsymbol{\delta} \neq \mathbf{0}$, where χ^2_2 is a Chi-square distribution with two degrees of freedom. The power is ap-

proximately $P_w(\hat{\boldsymbol{\delta}}, \boldsymbol{\delta}^*) = 1 - \Phi_{\chi^2_2}(\chi^2_{2,\alpha} - \boldsymbol{\delta}^{*\top}\hat{\mathbf{v}}(\hat{\boldsymbol{\delta}})^{-1}\boldsymbol{\delta}^*)$, where $\Phi_{\chi^2_2}$ is the CDF for a χ^2_2 distributed random variable and $\chi^2_{2,\alpha}$ is the corresponding α th upper percentile.

The power function $P_w(\hat{\delta}, \delta^*)$ for a given δ^*, or simply $P_w(\hat{\delta})$, is strict decreasing function of the variance $\hat{\mathbf{v}}(\hat{\delta})$. Thus, the results regarding variances in Theorems 4.1-4.3 can be directly translated in terms of power.

4.3 A simulation study

In the simulation study, we considered a continuous E and a bi-allelic G with genotype of AA, Aa and aa. In order to study the role of GE independence (assumption 4.1) and homogeneity in MAFs across cohorts (assumption 4.2), we considered $P(G, E)$ under four different settings, which reflect (a) both assumptions 4.1 and 4.2 hold; (b) assumption 4.1 holds but not 4.2; (c) assumption 4.2 holds but not 1; (d) neither assumption 4.1 nor 2 holds. To study the role of covariate heterogeneity in E, we considered both cases that wss is greater or smaller than bss for a fixed value of tss. The details of generating the gene environment data pair (G_{ki}, E_{ki}) for the i-th subject in the k-th study are described in Appendix B.3. Given (G_{ki}, E_{ki}), we generated the continuous trait Y_{ki} under the IPD model (4.1), where the study specific intercepts were sampled from $\alpha_k \overset{iid}{\sim} U(1.3, 1.5)$, the true effect sizes $(\beta^*_E, \beta^*_G, \delta^*)$ are determined such that E, G and GEI explain 10%, 1% and 0-1% of the total variation in Y respectively, in terms of partial R^2. The random residuals follow a $N(0, \sigma^2_k)$ distribution. The choice of σ^2_k leads to a marginal distribution of $Y \sim N(1.4, 0.4^2)$. The choice of $U(1.3, 1.5)$ and $N(1.4, 0.4^2)$ are motivated by the distribution of HDL-C ($mmol/l$) in our T2D data set. We generated $K = 20$ studies with different sample sizes involving a total of $N = 10,000$ participants ($n_k = 200$, for $k = 1, ..., 6$; $n_k = 400$, for $k = 7, ..., 11$; $n_k = 500$, for $k = 12, ..., 17$; $n_{18} = 800$; $n_{19} = 1000$; $n_{20} = 2000$).

We calculated $\hat{\delta}$ and $\hat{\mathbf{v}}(\hat{\delta})$ corresponding to each proposed estimator, including $\hat{\delta}^{\mathrm{IPD}}$, $\hat{\delta}^{\mathrm{UIVW}}$, $\hat{\delta}^{\mathrm{REM}}$, $\hat{\delta}^{\mathrm{MIVW}}$, $\hat{\delta}^{\mathrm{MR}}$, $\hat{\delta}^{\mathrm{TS}}$ and $\hat{\delta}^{\mathrm{AWE}}$. We carried out $R = 1,000$ replications under each setting, and summarized the results in terms of relative bias (RB $= (\frac{1}{R}\sum_{r=1}^{R}\hat{\delta}_{(r)} - \delta^*)/\delta^* \times 100\%$), mean of model based variance $\hat{\mathbf{v}}(\hat{\delta})$ (MV $= \frac{1}{R}\sum_{r=1}^{R}\hat{\mathbf{v}}(\hat{\delta}_{(r)})$), empirical variance (EV $= \frac{1}{R-1}\sum_{r=1}^{R}(\hat{\delta}_{(r)} - \overline{\hat{\delta}}_{(r)})^2$), mean squared error (MSE $= \frac{1}{R}\sum_{r=1}^{R}(\hat{\delta}_{(r)} - \delta^*)^2$) and power (the proportion of simulations that reject the null hypothesis using the Wald test for $H_0 : \delta = 0$). When the data is generated under the null, this proportion reduces to an empirical estimate of the Type-I error.

Lack of common set of confounders to adjust in each study: We then considered the situation where there exists additional explanatory variables $\boldsymbol{Z}$ in the true IPD model $Y_{ki} = \alpha_k + \beta_G G_{ki} + \beta_E E_{ki} + \delta G_{ki} E_{ki} + \boldsymbol{\beta}_Z^\top \boldsymbol{Z}_{ki} + \epsilon_{ki}$. We consider $\boldsymbol{Z} = (Z^1, Z^2, Z^3)$ mimicking typical covariates like (age, gender, race). In particular, age (Z^1) is continuous and associated with E, gender (Z^2) is binary and independent of both (G, E), race (Z^3) is a 3-level categorical variable and associated with both (G, E), and the true effect size $\boldsymbol{\beta}_Z^*$ is determined such that the Type-III partial R^2 corresponding to (Z^1, Z^2, Z^3) is $(2\%, 1\%, 1\%)$ respectively. Let $\boldsymbol{Z}_k$ be the set of covariates for the k-th study. We consider an analysis where $\boldsymbol{Z}_k$ is only partially available from individual studies, and refer to this situation as 'lack of common set of confounders to adjust'. In particular, we consider $\boldsymbol{Z}_k = (Z_k^1, Z_k^2, Z_k^3)$ for $k = 1, 2, 3$; $\boldsymbol{Z}_k = (Z_k^1, Z_k^2)$ for $k = 4, 5, 6$; $\boldsymbol{Z}_k = (Z_k^1, Z_k^3)$ for $k = 7, 8, 9$; $\boldsymbol{Z}_k = (Z_k^2, Z_k^3)$ for $k = 10, 11, 12$; $\boldsymbol{Z}_k = Z_k^1$ for $k = 13, 14$; $\boldsymbol{Z}_k = Z_k^2$ for $k = 15, 16$; $\boldsymbol{Z}_k = Z_k^3$ for $k = 17, 18$; No $\boldsymbol{Z}_k$ for $k = 19, 20$. For IPD analysis without any imputation of covariates, one can only obtain an IPD estimator based on the common subset of variables available across all studies, which reduces to an unadjusted model in our setting. We considered it as a naive IPD estimator ($\hat{\delta}^{\mathrm{NIPD}}$), and compared it to the true gold standard IPD estimator ($\hat{\delta}^{\mathrm{IPD}}$) as well as to the other estimators. For the meta-

analysis, we obtained $\hat{\delta}^{\mathrm{UIVW}}$, $\hat{\delta}^{\mathrm{MIVW}}$ and $\hat{\delta}^{\mathrm{REM}}$ from the k-th study model adjusted for $\boldsymbol{Z}_k$, for $k = 1, ..., K$. For MR, we adjusted $\boldsymbol{Z}_k$ at the first stage in the marginal genetic association model, and regressed the MLEs of adjusted effects of G on m_k.

Non-linear GEI model: We consider a non-linear GEI model where the phenotype-genotype association parameter $\beta_G(E)$ varies with E through a sigmoid function $\beta_G(E) = 2\exp(E-50)/\{1+\exp(E-50)\}+2$, as shown in Figure 4.1. In this case, $\beta_G(E)$ changes at different rates on different ranges of E (sharper around the mean value of E but relatively flat at more extreme values of E), which leads to non-linear interaction. In Figure 4.1, most studies only contribute to a restricted range of E, leading to heterogeneity of individual interaction estimates across studies. In this case, meta-analysis with a misspecified linear interaction model might fail to detect interaction. In the simulation study, we generated $K = 20$ studies, where 4 studies have relatively larger within study variability (studies 5, 10, 11, 15 in Figures 4.1 and 4.2) as compared to the other 16 studies. The complete description of n_k, m_k and σ_{Ek} for the 20 studies are given in Figures 4.1 and 4.2. We generated Y through the non-linear interaction model $Y_{ki} = \alpha_k + \beta_G(E_{ki})G_{ki} + \varepsilon_{ki}$, where $\varepsilon_{ki} \overset{iid}{\sim} N(0, \sigma_k^2)$. The within study relationship $\beta_G(E)$ are substantially different across studies. The effect heterogeneity and non-linearity might influence the relative performance of the proposed methods where a linear form of interaction is assumed. Therefore, we evaluated the robustness of the proposed meta-analysis estimators under this non-linear GEI model.

Simulation results

Comparison of methods: The relative performances of the methods are very similar across all susceptibility models and under all four settings, among which we only present the most general setting (d), where the data are generated without either assumption 4.1

or 2. In the main text, we compare the proposed methods in terms of power using the Wald test. In Figure 4.3, we observed three groups among the proposed methods: IPD, UIVW, REM, MIVW, MR, TS and AWE. Group 1: IPD, MIVW and AWE; group 2: UIVW, REM; group 3: MR. As expected, group 1 has the most powerful tests, which is consistent with LZ and our analytical results; group 2 is more powerful than group 3 if $Q_E < 1$, and vice versa. TS performs similarly as the better group between groups 2 and 3. $P_w(\hat{\delta}^{\mathrm{UIVW}})$ is slightly greater than $P_w(\hat{\delta}^{\mathrm{REM}})$ since the underlying model is FEM. The empirical estimates of Type-I error are close to the true 0.05 level for all tests under all three susceptibility models and all four settings. Power curves under settings (a)-(c) are given in Appendix A.3, where similar results were shown. Additional simulation results for RB, MV, EV and MSE are given in the Appendix A.3.

Covariate heterogeneity in E: We observed that the ARE between $\hat{\delta}^{\mathrm{UIVW}}$ ($\hat{\delta}^{\mathrm{MR}}$) and $\hat{\delta}^{\mathrm{IPD}}$ can be well characterized in terms of wss/tss (bss/tss) respectively. We found that $\hat{\delta}^{\mathrm{UIVW}}$ was more efficient than $\hat{\delta}^{\mathrm{MR}}$ if $wss > bss$, and vice versa. The precision trade-off is captured well by the adaptively determined weights in $\hat{\delta}^{\mathrm{AWE}}$. We observed that $\hat{\delta}^{\mathrm{AWE}}$ is more efficient than the usual meta-analytic estimators $\hat{\delta}^{\mathrm{UIVW}}$, $\hat{\delta}^{\mathrm{REM}}$, $\hat{\delta}^{\mathrm{MR}}$ or $\hat{\delta}^{\mathrm{TS}}$, and had almost the same efficiency as $\hat{\delta}^{\mathrm{IPD}}$ and $\hat{\delta}^{\mathrm{MIVW}}$ under all three susceptibility models and all four settings. The findings are consistent with LZ and our analytical results in Theorems 4.2 and 4.3.

Gene-environment independence: Comparing settings (a) to (c) (or alternatively comparing settings (b) to (d)), where the only difference is the dependence between G and E, we observed no substantial difference in RB for all of the proposed estimators, including the potentially biased estimators $\hat{\delta}^{\mathrm{MR}}$ and $\hat{\delta}^{\mathrm{AWE}}$. When assumption 4.1 is relaxed (settings (c) and (d)), the magnitude of bias of $\hat{\delta}^{\mathrm{MR}}$ (up to $\pm 3\%$) and $\hat{\delta}^{\mathrm{AWE}}$ (up to $\pm 2\%$) is not yet to a level of practical concern compared to the Monte Carlo error (up to $\pm 3\%$ even

for the unbiased estimators). For variance, we did not observe precision gain by making the GE independence assumption as expected. When assumption 4.1 is relaxed, results in Theorem 4.2 hold numerically for all three genetic susceptibility models.

Homogeneity in allele frequencies across cohorts: Comparing settings (a) to (b) (or alternatively comparing settings (c) to (d)), we did not observe precision gain when the MAF is homogeneous across studies (settings (a) and (c)). When assumption 4.2 is relaxed, results in Theorem 4.3 hold numerically for all three genetic susceptibility models.

Lack of common set of confounders to adjust in each study: Figure 4.4 shows the power curves under this situation without either assumption 4.1 or 2. Compared to the basic setting without covariate adjustment (Figure 4.3), there is no substantial difference in the relative performances of testing among these methods. We observed that the GEI estimate $\hat{\delta}$ and variance $\hat{\mathbf{v}}(\hat{\delta})$ was fairly unchanged, though the main effects of $\hat{\beta}_G$ and $\hat{\beta}_E$ were substantially influenced under this situation. VanderWeele et al. (2012) also showed similar results that, under GE independence, there is no effect of unmeasured environmental confounding on the GEI parameter; and that if G and E are dependent, the environmental confounding needs to be very strong to incur substantial bias in GEI. Power curves under settings (a)-(c) are given in Appendix A.3, where similar results were shown.

Misspecification of the genetic susceptibility model: We examined the power under misspecified susceptibility models (dominant/additive), where the true generating model is co-dominant. When $\delta^{AA} = 1.5\delta^{Aa}$ (we accordingly choose $\beta_G^{AA} = 1.5\beta_G^{Aa}$), i.e., the second copy of A has an effect size between the two assumed in dominant ($\delta^{AA} = \delta^{Aa}$) and additive ($\delta^{AA} = 2\delta^{Aa}$) models, there is no substantial difference of power between the misspecified dominant/additive model and co-dominant model (shown in Appendix A.3), because the misspecification is not strong and the fitted dominant or additive models used one less parameter. When $\delta^{AA} = -\delta^{Aa}$ (we accordingly choose $\beta_G^{AA} = -\beta_G^{Aa}$), i.e., the

second copy of A has a reverse effect, the fitted dominant or additive models had much less power than the co-dominant model (shown in Figures 4.5). Thus, it could happen that the co-dominant model has more power compared to other simpler models, though it uses two additional parameters for capturing GEI.

Non-linear GEI model: When the IPD were generated under the non-linear GEI model, the power to detect GEI from individual studies were very low (< 0.25), except study 10 where the sample size n_{10}, effect size (depends on E) and variance $\sigma^2_{E_{10}}$ are all relatively greater than the other studies (Figure 4.2). In Table 4.2, $P_w(\hat{\delta}^{\mathrm{IPD}})$, $P_w(\hat{\delta}^{\mathrm{MIVW}})$ and $P_w(\hat{\delta}^{\mathrm{AWE}})$ show the highest powers. $P_w(\hat{\delta}^{\mathrm{MIVW}})$ is close to $P_w(\hat{\delta}^{\mathrm{IPD}})$ because the model based standard errors of $\hat{\delta}^{\mathrm{IPD}}$ and $\hat{\delta}^{\mathrm{MIVW}}$ are asymptotically the same. Because most of the 20 studies were unable to represent the true non-linear GEI, especially those with very short range of E, the non-linearity of GEI lead $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{REM}}$ to be low. In this particular example, we observed that $P_w(\hat{\delta}^{\mathrm{MR}})$ is greater than $P_w(\hat{\delta}^{\mathrm{UIVW}})$. Instead of choosing alternatively between $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$, we can use $\hat{\delta}^{\mathrm{AWE}}$ as the default meta-analytic estimator. The relative performance of $\hat{\delta}^{\mathrm{AWE}}$ is close to $\hat{\delta}^{\mathrm{IPD}}$. This is a practically noteworthy finding as a linear interaction model is typically the initial screening tool, and the AWE is able to pick up signals under model misspecification that univariate meta-analysis methods can not.

Alternatively, one can consider a stratified analysis where the 3 strata consist of study k with m_k falling into the three intervals $m_k \leq 48$, $48 < m_k < 52$ and $52 \leq m_k$ respectively such that the value of GEI is close within each stratum (refer to Figure 4.1). Then $\hat{\delta}^{\mathrm{UIVW}}$ was used for meta-analysis of GEI within each stratum. We observed a power of 0.80 for stratum $(48, 52)$, and only around 0.15 for the other two strata. The Wald test of $\hat{\delta}^{\mathrm{UIVW}}$ were still powerful around the mean value of E but lack of power at more extreme values of E.

4.4 Application: Data analysis for a set of studies investigating type 2 diabetes

The proposed methods were applied to a set of studies investigating T2D, including 8 European cohorts: D2D2007, DIAGEN, DPS, FUSION FSIGT, FUSION S2, HUNT, METSIM and TROMSO. A number of SNPs in the *FTO* gene region (16q12.2) have previously been identified to be associated with T2D and BMI in the DIAGRAM consortium (Zeggini et al. (2008), Voight et al. (2010)). Variants at *FTO* are known to influence T2D predisposition through an effect on BMI (Freathy et al. (2008), Voight et al. (2010)). Age, BMI and gender are all known risk factors for T2D and a T2D related quantitative trait HDL-C (Scott et al. (2012), Morris et al. (2012)). In this chapter, we investigated whether SNPs in *FTO* gene modifies the effect of environmental factors (e.g. age and BMI) on HDL-C. No association between SNPs in *FTO* and HDL-C or SNP$\times$BMI interaction on HDL-C has ever been noted.

Among the 8 cohorts, the T2D patients were identified by the glucose tolerance category (fasting glucose$\geq 7.0\ mmol/l$ or two-hour glucose$\geq 11.1\ mmol/l$). We genotyped all the T2D patients and portions of non-T2D participants under budget allowance. Thus, T2D patients were over sampled in the genotyped data set as compared to the overall cohorts. The descriptive summary statistics for the genotyped data sets from the 8 cohorts are shown in Table 4.3. We have a total of $N = 11,729$ genotyped participants who have HDL-C levels available, with sample sizes n_k ranging between 172 and 2,730. Since the SNPs in the *FTO* gene we examined (10 strongest SNPs associated with T2D/obesity/BMI that listed on the National Human Genome Research Institute (NHGRI) GWAS catalog) are in high linkage disequilibrium and show very similar results. Thus, we present our results for one representative SNP, rs1121980, only. The SNP follows Hardy-Weinberg equilibrium (HWE), and no imputation was needed as the missing genotype proportion

is $< 0.1\%$. The MAF of rs1121980 ranges 0.40-0.49 across cohorts, as an evidence for supporting assumption 4.2. In Table 4.3, the mean age ranges from 56-69 (years) except FUSION FSIGT cohort (mean age=39). FUSION FSIGT cohort appears younger because either spouse or offspring of T2D sibpairs were selected as a follow-up to the FUSION study. The mean BMI ranges from 26-28 (kg/m^2) except the DPS cohort (mean BMI=31), because the DPS cohort has an inclusion criterion requiring all subjects have BMI> 25 at baseline. Thus, the covariate heterogeneity of E are small, with $BSS_{age}/TSS_{age} = 14\%$ and $BSS_{BMI}/TSS_{BMI} = 2\%$ respectively. The two 'outlier' cohorts both have only very small sample sizes compared to the other studies, so their influence on UIVW and MIVW should be small. However, the influence on MR could be very substantial due to a small number of studies.

Analysis Model: The IPD model we fitted is given by

$$\begin{aligned} \log(\text{HDL-C}_{ki}) &= \alpha_k + \beta_G G_{ki} + \delta G_{ki} \times E_{ki} \\ &+ \beta_{age} age_{ki} + \beta_{BMI} BMI_{ki} + \beta_{gender} gender_{ki} + \epsilon_{ki}, \end{aligned} \tag{4.4}$$

for $k = 1, ..., 8$; $i = 1, ..., n_k$. In model (4.4), SNP rs1121980 was used for G; BMI or age was considered as E in two separate models for rs1121980×BMI and rs1121980×age interactions respectively; HDL-C was log-transformed in order to reduce the skewness of its distribution. The proposed methods, including IPD, UIVW, REM, MIVW (MIVW2'), MR, TS and AWE (AWE2'), were implemented and compared. GE independence does not appear to hold for rs1121980×BMI analysis (Spearman correlations across studies were reported in Table 4.3). This is expected as *FTO* is an obesity related gene. GE independence appears to hold for rs1121980×age (Table 4.3). We also considered adjusting for the T2D status in model (4.4). Since the results are very similar, we only show the results corresponding to model (4.4) for demonstration purpose.

Results: Figure 4.6 shows the forest plots of estimated GEI from individual cohorts ($\hat{\delta}_k$) as well as the combined estimates using joint or meta-analysis. The corresponding numerical results were provided in Table 4.4. There was no evidence of effect heterogeneity for both rs1121980×BMI ($P = 0.59$) and rs1121980×age ($P = 0.81$) interaction based on Cochran's Q test, and UIVW and REM showed similar results. We observed that, under rs1121980×BMI model, $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}})/\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})$ is 0.94; $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}})/\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MIVW}})$ is 0.95; $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}})/\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{AWE}})$ is 0.98. Under rs1121980×age model, $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}})/\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})$ is 0.75; $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}})/\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MIVW}})$ is 0.90; $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}})/\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{AWE}})$ is 0.98. These ratios were potentially determined by the covariate heterogeneity of E. All these meta-analytical methods UIVW, REM, MIVW (MIVW2′) and AWE (AWE2′) showed very similar results as IPD, especially for rs1121980×BMI interaction. The marginal SNP effects of rs1121980 against mean covariate values of age and BMI across cohorts are shown in Appendix A.3. MR was very sensitive to outliers when the number of cohorts was small (K=8), and showed quite different results from the other methods due to outliers (Table 4.4). MR also lacked efficiency for a small K and small ratio BSS/WSS. $\hat{\delta}^{\mathrm{AWE}}$ was robust to the bias from $\hat{\delta}^{\mathrm{MR}}$ since it only assigned weight $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})^{-1}/\{\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})^{-1}\} = 0.04$ on $\hat{\delta}^{\mathrm{MR}}$. This is further evidence that $\hat{\delta}^{\mathrm{AWE}}$ can data adaptively shrink to the 'better' estimator. Moreover, $\hat{\delta}^{\mathrm{AWE}}$ showed almost full efficiency as compared to $\hat{\delta}^{\mathrm{IPD}}$.

In the model that drops the interaction term in (4.4), no significant marginal effect of rs1121980 was found (at 0.05 level) after adjusting for the risk factors of age, gender and BMI. It is expected since SNPs in *FTO* are known to influence T2D predisposition through the effect on BMI. These risk factors themselves were strongly significant. Mean HDL-C level decreased by: 0.10% (95% CI: (0.05, 0.14)%) for 1 year increase in age; 1.65% (95% CI: (1.56, 1.75)%) for a 1 kg/m^2 increase in BMI; 20.0% (95% CI: (18.7, 21.1)%) from female to male, conditional on the other risk factors in the model.

In our interaction model (4.4), positive rs1121980×BMI interactions were found under all proposed methods (except MR) in Table 4.4, with P-values range from 0.005 to 0.007 (for additive model). In particular, the estimates obtained from model (4.4) when converted in terms of percentage change in actual HDL-C levels, indicated that: with 1 kg/m^2 increase in BMI, (1) under additive model, on average HDL-C level decreased by 1.73% (95% CI: (1.57, 1.90)%) given rs1121980=GG, by 1.54% (95% CI: (1.44, 1.64)%) given rs1121980=AG or GA; and by 1.35% (95% CI: (1.17, 1.53)%) given rs1121980=AA; (2) under co-dominant model, HDL-C level decreased by 1.70% (95% CI: (1.51, 1.88)%) given rs1121980=GG, by 1.58% (95% CI: (1.44, 1.72)%) given rs1121980=AG or GA; and by 1.31% (95% CI: (1.10, 1.53)%) given rs1121980=AA. The results under additive and co-dominant models were very close. The trend of the effects of BMI among the three groups defined by rs1121980 indicated that the presence of minor allele A in rs1121980 attenuated the negative association between BMI and HDL-C. We did not find similar rs1121980×BMI interaction effect on other lipid traits related to T2D, including low-density lipoprotein cholesterol (LDL-C), total cholesterol and LDL-C/HDL-C ratio (P-value = 0.08).

4.5 Discussion

In this chapter, we proposed and compared a set of meta-analysis approaches for analyzing GEI. We showed the proposed AWE, as a combination of meta-analysis and meta-regression estimators, performed better than alternatively choosing between the two estimators in terms of precision and power. We showed that the precision trade-off between the two components in AWE depends on the covariate heterogeneity of E, and that the weights in AWE can adaptively capture this trade-off. The resulting AWE retains full efficiency of the 'gold standard' joint analysis using IPD under certain natural assumptions.

We suggest to use AWE as a default choice for the meta-analysis of GEI based on summary data. We studied several key features that could potentially influence the efficiency and power for meta-analysis of GEI, and hoped to provide useful insight and guidelines for such studies. The features included: (1) departures from GE independence; (2) heterogeneity in MAFs across cohorts; (3) lack of a common set of confounders to adjust for in individual studies; (4) misspecification of the genetic susceptibility model (dominant/co-dominant/additive); and (5) the presence of a non-linear form of interaction. Under all the above situations, we found the relative performance of AWE is close to IPD estimator. We especially would like to point out the simulation findings under the non-linear interaction model setting, where standard meta-analytical technique failed and the AWE was able to capture the lost efficiency based on the summary data. We also reported some evidence for GEI between rs1121980 and BMI on HDL-C levels.

We have mainly focused on quantitative traits with an underlying FEM. The potential limitation is that the results might not be able to directly translate to dichotomous traits under a case-control design, where assuming GE independence leads to huge gain in efficiency (Piegorsch et al. (1994), Umbach and Weinberg (1997), Chatterjee and Carroll (2005)). We plan to extend our methods using a retrospective likelihood framework under a case-control design. Investigating the results under a truly random effects meta-analysis model is another possible extension to our work.

Figure 4.1: Non-linear GEI model: the (red) sigmoid curve shows the true relationship between Y-G association and E, $\beta_G(E) = 2\exp(E-50)/\{1+\exp(E-50)\}+2$; the boxplots show the covariate heterogeneity of E across studies where the dots show the corresponding means.

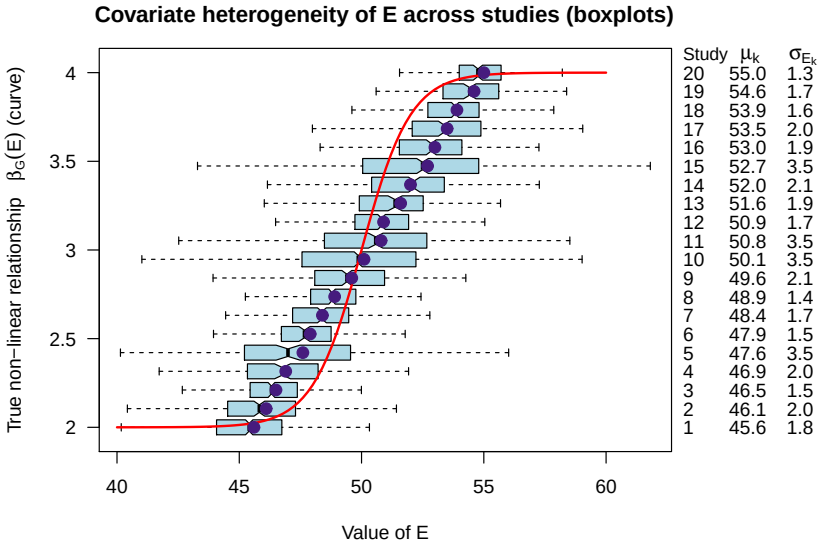


Figure 4.2: Non-linear GEI model: the barchart shows the power to detect GEI across individual studies; the (green) curve shows the value of the true non-linear GEI; the top panel shows the sample sizes n_k and the within study standard deviations σ_{E_k} of E, where the four studies with relatively greater σ_{E_k} are highlighted (in red).

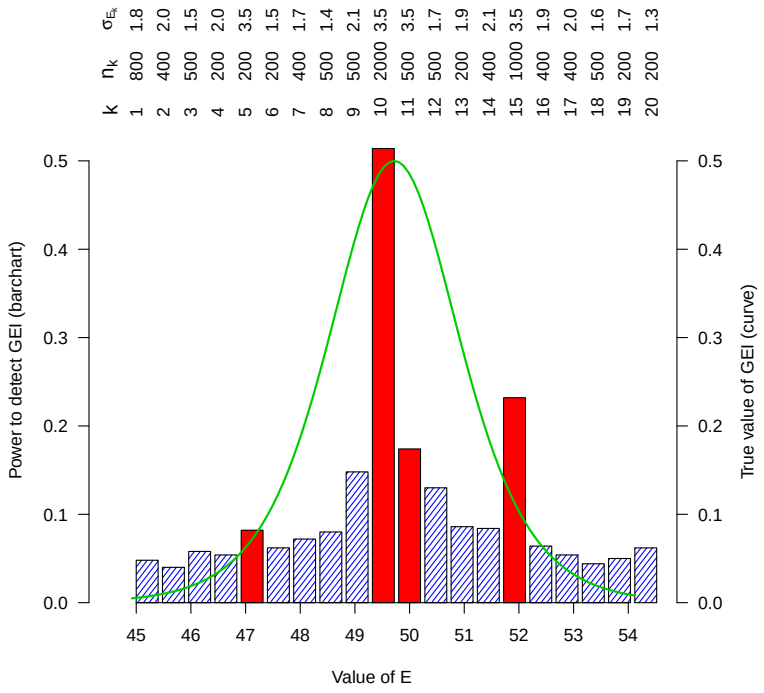


Figure 4.3: Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study, where no assumption of G-E independence or homogeneity in allele frequencies is assumed.

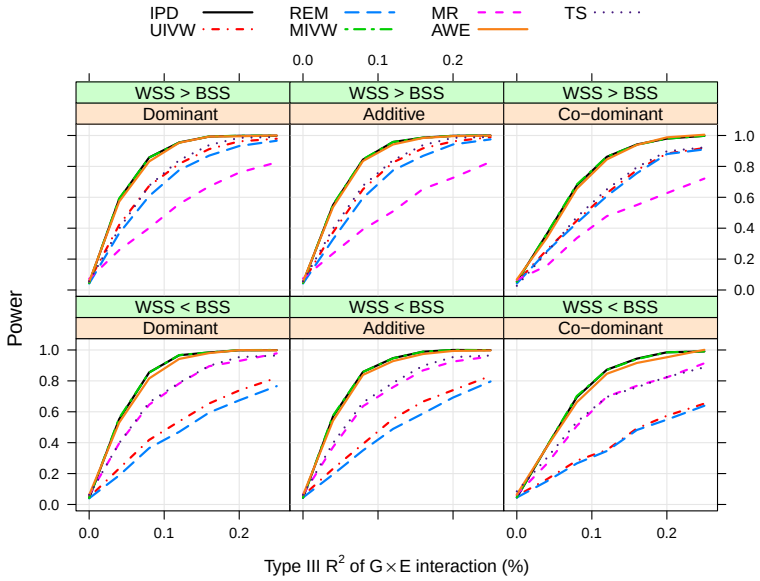


Figure 4.4: Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study (for the situation of lack of common set of confounders to adjust), where no assumption of gene-environment independence or homogeneity in allele frequencies is assumed.

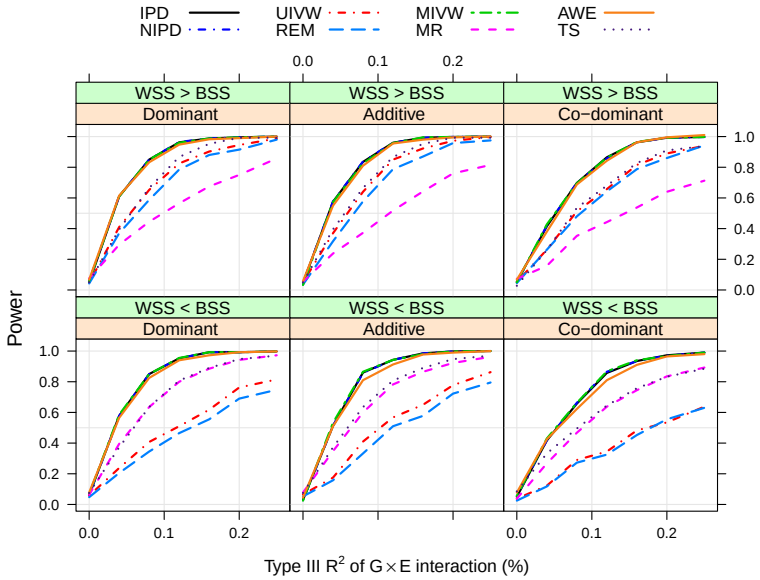


Figure 4.5: Power curves under misspecified susceptibility models (dominant/additive), where the generating co-dominant model has δ^{AA}=$-\delta^{Aa}$, and no assumption of gene-environment independence or homogeneity in allele frequencies is assumed.

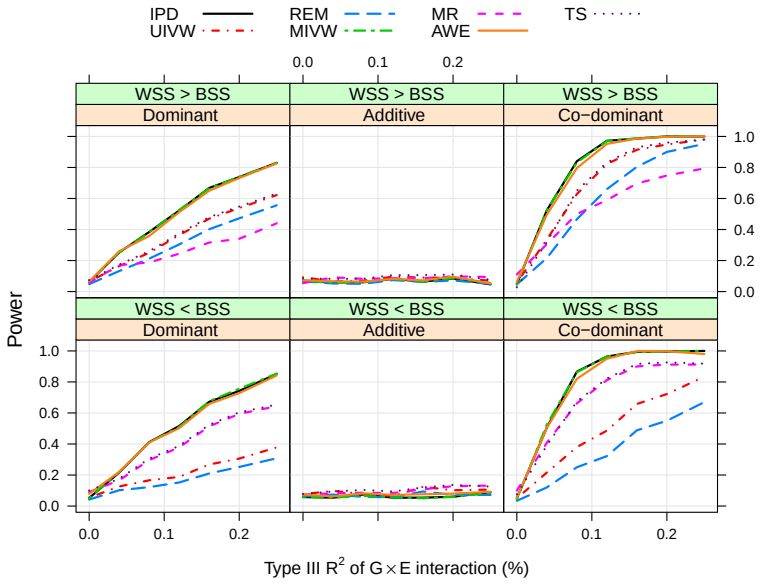


Figure 4.6: Forest plots showing the estimated gene-environment interactions (under additive model of rs1121980) across the 8 European cohorts, as well as the combined estimates through meta-analysis. [IPD: individual patient data; UIVW: univariate inverse-variance weighted estimator; REM: random effect model; MIVW: multivariate inverse-variance weighted estimator pooling $(\beta_G, \beta_E, \delta)$; MIVW2′: MIVW estimator pooling (β'_E, δ') under a centered model; AWE: adaptively weighted estimator combining UIVW and Meta-regression; AWE2′: AWE combining MIVW2′ and Meta-regression.]

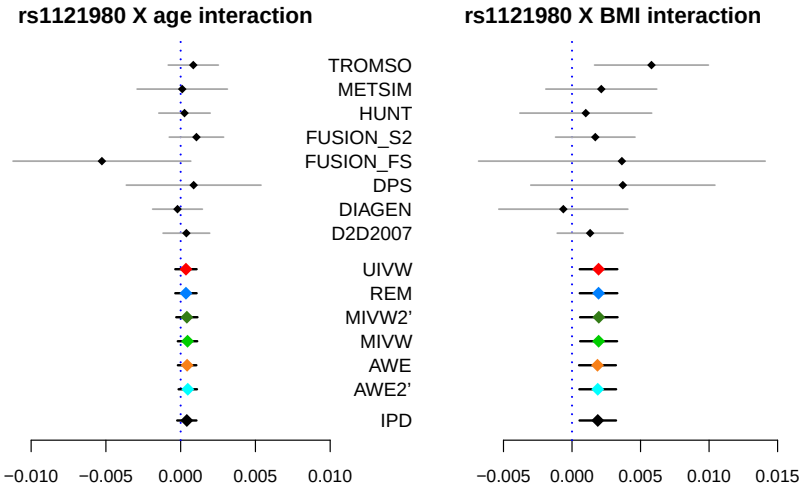


Table 4.1: Operating characteristics for the meta-analysis of GEI. [IPD: individual patient data analysis; UIVW: univariate inverse-variance weighted estimator; MIVW: multivariate inverse-variance weighted estimator; MR: Meta-regression; AWE: adaptively weighted estimator.]

Methods	Data shared	Bias	ARE[a]
IPD	individual level data	unbiased	1
UIVW	$\hat{\delta}_k$, $\hat{\mathbf{v}}(\hat{\delta}_k)$	unbiased	wss/tss
MIVW	$\hat{\boldsymbol{\beta}}_k$, $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)$	unbiased	1
MR	$\hat{\lambda}_k$, $\hat{\mathbf{v}}(\hat{\lambda}_k)$ and m_k	unbiased under assumption 4.1 ecological bias in general	bss/tss
AWE	$\hat{\delta}_k$, $\hat{\mathbf{v}}(\hat{\delta}_k)$, $\hat{\lambda}_k$, $\hat{\mathbf{v}}(\hat{\lambda}_k)$ and m_k	unbiased under assumption 4.1 bias adaptively controlled	1

[a] ARE: asymptotic relative efficiency as compared to $\hat{\delta}^{\mathrm{IPD}}$ under assumptions 1 and 2.

Table 4.2: Comparison of methods in terms of estimate, variance and power, under a simulation study of non-linear GEI. [IPD: individual patient data; UIVW: univariate inverse-variance weighted estimator; REM: random effect model; MIVW: multivariate inverse-variance weighted estimator pooling $(\beta_G, \beta_E, \delta)$; MIVW2′: MIVW estimator pooling (β'_E, δ') under a centered model; MR: Meta-regression; AWE: adaptively weighted estimator combining UIVW and MR; AWE2′: AWE combining MIVW2′ and MR; TS: two-stage approach.]

Methods	Estimate	SE[a]	Power
IPD	0.21	0.045	0.98
UIVW	0.18	0.070	0.69
REM	0.19	0.076	0.67
MIVW2′	0.18	0.070	0.73
MIVW	0.21	0.045	0.98
MR	0.23	0.060	0.82
AWE	0.21	0.045	0.98
AWE2′	0.21	0.045	0.98
TS			0.85

[a] SE: standard error.

Table 4.3: Summary statistics for the genotyped data set from the 8 European cohorts.

Cohorts	N	HDL-C ($mmol/l$) Mean (SD)	rs1121980 MAF[a]	Age (year) Mean (SD[a])	BMI (kg/m^2) Mean (SD)	Gender Female (%)	SNP*Age Corr (P)[b]	SNP*BMI Corr (P)
D2D2007	2693	1.44 (0.35)	0.41	59.9 (8.4)	27.5 (4.8)	52	-0.03 (0.16)	0.03 (0.19)
DIAGEN	1510	1.45 (0.47)	0.46	63.3 (14.3)	27.9 (5.2)	55	-0.01 (0.76)	0.03 (0.24)
DPS	433	1.22 (0.29)	0.44	55.1 (7.1)	31.3 (4.6)	68	-0.02 (0.69)	0.16 (<.01)
FUSION-FS	172	1.29 (0.32)	0.43	38.6 (10.9)	26.2 (4.9)	55	0.04 (0.56)	0.23 (<.01)
FUSION-S2	2730	1.45 (0.41)	0.40	57.2 (8.4)	27.9 (5.1)	44	-0.02 (0.23)	0.06 (<.01)
HUNT	1324	1.26 (0.38)	0.47	67.2 (13.1)	28.0 (4.4)	48	<.01 (0.94)	0.06 (0.03)
METSIM	1456	1.42 (0.40)	0.44	56.3 (6.6)	27.9 (4.7)	0	-0.05 (0.08)	0.03 (0.32)
TROMSO	1411	1.43 (0.42)	0.49	59.9 (12.5)	27.6 (4.7)	50	<.01 (0.91)	0.04 (0.15)
Entire study	11729	1.41 (0.40)	0.43	59.6 (11.1)	27.9 (4.9)	44	<.01 (0.97)	0.05 (<.01)

[a] SD: standard deviation; BMI: body mass index; MAF: minor allele frequency;
[b] Corr(P): Spearman correlation between SNP rs1121980 and environmental factor with corresponding P-value.

Table 4.4: IPD/Meta-analysis results of GEI for the T2D study, where the log transformed HDL-C level was regressed on SNP, age, BMI, gender, cohorts, and either SNP×BMI or SNP×age interaction in the IPD model. Estimates, SEs and CIs have been multiplied by 1000.

Methods[a]	rs1121980 (additive) × BMI			P-value*	
	Estimate	SE[b]	95% CI[b]	Additive	Co-dominant
IPD	1.880	0.679	(0.548, 3.212)	0.006**	0.013*
UIVW	1.936	0.700	(0.565, 3.308)	0.006*	0.018*
REM	1.936	0.700	(0.565, 3.308)	0.006*	0.018*
MIVW2′	1.955	0.698	(0.587, 3.324)	0.005*	0.017*
MIVW	1.957	0.698	(0.589, 3.325)	0.005*	0.017*
MR	-0.062	3.484	(-6.890, 6.767)	0.986	0.630
AWE	1.859	0.686	(0.514, 3.204)	0.007*	0.016*
AWE2′	1.877	0.685	(0.536, 3.219)	0.006*	0.015*
	rs1121980 (additive) × age			Additive	Co-dominant
IPD	0.381	0.314	(-0.235, 0.998)	0.225	0.324⋆
UIVW	0.354	0.364	(-0.360, 1.068)	0.331	0.372
REM	0.354	0.364	(-0.360, 1.068)	0.331	0.372
MIVW2′	0.415	0.362	(-0.295, 1.124)	0.252	0.387
MIVW	0.458	0.332	(-0.193, 1.109)	0.168	0.397
MR	0.666	0.633	(-0.575, 1.906)	0.293	0.471
AWE	0.432	0.316	(-0.187, 1.050)	0.172	0.191
AWE2′	0.477	0.314	(-0.140, 1.093)	0.129	0.184

[a] IPD: individual patient data; UIVW: univariate inverse-variance weighted estimator; REM: random effect model; MIVW: multivariate inverse-variance weighted estimator pooling $(\beta_G, \beta_E, \delta)$; MIVW2′: MIVW estimator pooling (β'_E, δ') under a centered model; MR: Meta-regression; AWE: adaptively weighted estimator combining UIVW and MR; AWE2′: AWE combining MIVW2′ and MR.
[b] SE: standard error; CI: confidence interval.
* indicating significance at $\alpha = 0.05$ level.
⋆ indicating whether additive or co-dominant model has smaller AIC under the IPD model.

CHAPTER V

Meta-analysis of gene-environment interaction on dichotomous traits under case-control studies

5.1 Introduction

5.1.1 The role of G-E independence in case-control studies of G-E interaction

Gaining efficiency in studies of gene-environment interaction (GEI) by exploiting independence between the genetic (G) and environmental (E) factors under case-control sampling has been noted in multiple papers (Piegorsch et al. (1994); Umbach and Weinberg (1997); Chatterjee and Carroll (2005), CC from now on). Piegorsch et al. (1994) showed that one can estimate multiplicative GEI in logistic model with data from cases alone, provided that G and E are independent in the population and the disease is rare. Under such assumptions, the GEI parameter is obtained as the odds ratio between G and E among cases, which is more precise than that obtained from a case-control analysis using logistic regression. However, the case-only method can only make inference on the GEI parameter without getting the corresponding main effect simultaneously. Umbach and Weinberg (1997) showed that, with data available on both cases and controls, one can estimate the main effects and interaction, by fitting a suitably constrained log-linear model under the rare disease assumption and the G-E independence assumption, which has the same precision of the GEI parameter derived under the case-only method. However, methods in Piegorsch et al. (1994) and Umbach and Weinberg (1997) only work for

categorical E. CC proposed a semi-parametric maximum likelihood method to estimate all the logistic regression parameters under a retrospective likelihood with continuous E and possible covariate adjustment. Their method addressed many of the limitations of the existing methods as discussed above, e.g. the rare disease assumption, categorical E. They also considered the issue of population stratification where G-E independence assumption only holds conditional on the set of stratification variables. However, methods that use G-E independence assumption might produce severely biased estimates if the assumption is violated. Several studies have addressed this issue and proposed more robust strategies for testing GEI (Mukherjee and Chatterjee (2008), MC from now on; Mukherjee et al. (2008); Li and Conti (2009); Murcray et al. (2009)). MC proposed a solution to the bias versus efficiency dilemma, using a retrospective method that allows for uncertainty around the assumption of G-E independence. MC used the estimate of the uncertainty parameter in an empirical Bayes (EB) fashion to obtain a shrinkage estimator that 'shrinks' the maximum likelihood estimates (MLEs) of disease odds ratio parameters under G-E dependence to those under G-E independence, and showed how the shrinkage factor depends on these MLEs and their corresponding variances. The following is a detailed review.

5.1.2 Review of MC's empirical Bayes approach

Let D be the binary indicator of presence ($D = 1$) or absence ($D = 0$) of a disease. Let G, E and S be the genotype, environmental exposure and stratification factor (such as ethnicity). MC considered the following factorization of the retrospective likelihood for a case-control study,

$$L^R = P(G, E, S|D) = \frac{P(D|G, E, S)P(G|E, S)P(E, S)}{\sum_{G,E,S} P(D|G, E, S)P(G|E, S)P(E, S)}.$$

MC considered (1) a logistic disease incidence model $P(D|G, E, S) = H\{\gamma_0 + \gamma_G G + \gamma_E E + \gamma_{GE} GE + \gamma_S S\}$, where $H(u) = \{1 + \exp(-u)\}^{-1}$; (2) a logistic model $P(G =$

$1|E,S) = H\{\eta_0 + \theta E + \eta S\}$ for a binary genetic factor G; (3) the joint distribution for (E,S) to be nonparametric. In (2), θ is a measure of dependence between G and E conditional on S. Under G-E independence conditional on S, $P(G|E,S) = P(G|S)$. One can reduce the model for G to $P(G=1|S) = H\{\eta_0 + \eta S\}$. Throughout this chapter, we refer to the model with (without) G-E independence assumption as constrained (unconstrained) model respectively.

The EB estimator in MC is under a general framework, where one is interested in inference on a set of focus parameters $\boldsymbol{\beta}$ in the presence of prior information only on a set of nuisance parameters $\boldsymbol{\theta}$. For example, under the above formulation (1)-(3), $\boldsymbol{\beta} = (\gamma_0, \boldsymbol{\gamma}, \eta_0, \eta)$ and $\boldsymbol{\theta} = \theta$ (We use $\boldsymbol{\theta}$ instead of θ below for the general case). Denote the MLE for $\boldsymbol{\beta}$ under the unconstrained and constrained model as $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}^0$ respectively. Denote $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ as the profile MLE of $\boldsymbol{\beta}$ for fixed $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\beta}}^0 = \hat{\boldsymbol{\beta}}(\mathbf{0})$. Denote $\boldsymbol{\beta}(\boldsymbol{\theta})$ as the limiting value of $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$. Following the invariant property of MLE, we have $\boldsymbol{\beta}(\hat{\boldsymbol{\theta}}) \sim MVN(\boldsymbol{\beta}(\boldsymbol{\theta}), \boldsymbol{V}_{\beta(\hat{\theta})})$. We use the generic notation $\boldsymbol{V}_{(\cdot)}$ to denote the variance of $(\cdot)$. Assumes $\boldsymbol{\theta} \sim MVN(\mathbf{0}, \boldsymbol{A})$. Applying Taylor's expansion around $\boldsymbol{\theta} = \mathbf{0}$, MC used the linear approximation $\boldsymbol{\beta}(\boldsymbol{\theta}) \sim MVN(\boldsymbol{\beta}(\mathbf{0}), \Delta^\top \boldsymbol{A} \Delta)$, where $\Delta^\top = \partial \boldsymbol{\beta}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\mathbf{0}}$ is the Jacobian matrix evaluated at $\mathbf{0}$. For the Gaussian-Gaussian model, MC proposed an approximation to the Bayes estimator of $\boldsymbol{\beta}(\boldsymbol{\theta})$ for a fixed $\boldsymbol{A}$ as

$$\widehat{\boldsymbol{\beta}(\boldsymbol{\theta})} = \Delta^\top \boldsymbol{A} \Delta \{\boldsymbol{V}_{\beta(\hat{\theta})} + \Delta^\top \boldsymbol{A} \Delta\}^{-1} \boldsymbol{\beta}(\hat{\boldsymbol{\theta}}) + \boldsymbol{V}_{\beta(\hat{\theta})} \{\boldsymbol{V}_{\beta(\hat{\theta})} + \Delta^\top \boldsymbol{A} \Delta\}^{-1} \boldsymbol{\beta}(\mathbf{0}).$$

MC further used the profile MLE $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ for $\boldsymbol{\beta}(\boldsymbol{\theta})$, the estimated asymptotic variance $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}$ for $\boldsymbol{V}_{\beta(\hat{\theta})}$, $\hat{\Delta} = \partial \hat{\boldsymbol{\beta}}^\top(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\mathbf{0}}$ for Δ, and a conservative estimator $\hat{\boldsymbol{A}} = \hat{\boldsymbol{\theta}} \hat{\boldsymbol{\theta}}^\top$ for $\boldsymbol{A}$. Then the resulting ad hoc EB estimator is given by

$$\hat{\boldsymbol{\beta}}_{EB} = \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta} \{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}} + \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta}\}^{-1} \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}} \{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}} + \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta}\}^{-1} \hat{\boldsymbol{\beta}}^0.$$

The variance $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}_{EB}}$ of $\hat{\boldsymbol{\beta}}_{EB}$ can be estimated using the multivariate Taylor's expansion. A

Wald-type test can be constructed based on the statistic $\hat{\boldsymbol{\beta}}_{EB}^{\top}\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}_{EB}}^{-1}\hat{\boldsymbol{\beta}}_{EB}$.

Alternatively, one can write $\hat{\boldsymbol{\beta}}_{EB} = \hat{\boldsymbol{\beta}} + \boldsymbol{K}(\hat{\boldsymbol{\beta}}^0 - \hat{\boldsymbol{\beta}})$, where $\boldsymbol{K} = \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}}\{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\beta}}} + \hat{\Delta}^{\top}\hat{\boldsymbol{A}}\hat{\Delta}\}^{-1}$. The shrinkage factor $\boldsymbol{K}$ determines the amount of shrinkage between the unconstrained and constrained MLEs. When $\boldsymbol{K} \to \mathbf{1}$ (evidence of G-E independence), we have $\hat{\boldsymbol{\beta}}_{EB} = \hat{\boldsymbol{\beta}}^0$; when $\boldsymbol{K} \to \mathbf{0}$ (departure from G-E independence), we have $\hat{\boldsymbol{\beta}}_{EB} = \hat{\boldsymbol{\beta}}$. So MC's EB shrinkage estimator can relax the G-E independence assumption in a data-adaptive fashion. Further theoretical development regarding this shrinkage estimator is presented in Chen et al. (2009). It was noted that this EB estimate can trade off between bias and efficiency and provide increased power compared to a standard case-control analysis, with superior control of type 1 error when compared to a case-only analysis. This performance is noted for modest sample sizes, whereas asymptotically the EB estimator $\hat{\boldsymbol{\beta}}_{EB}$ converges to the unconstrained estimator $\hat{\boldsymbol{\beta}}$

As the GEIs detected so far only have small to modest effects, there are increasing demands for large sample sizes and collaboration across different study sites in order to perform a pooled or meta-analysis with high confidence and power. In this chapter, we consider possible extensions of EB type shrinkage estimators for a multiple-study setting. To handle this multiple-study problem, we particularly consider strategies to obtain a shrinkage factor in the EB estimator that can borrow strength across studies, under both individual patient data (IPD) analysis using individual level data and meta-analysis using study level summary statistics.

5.2 Methods

5.2.1 Model formulation under multiple-study setting

Suppose there are K independent studies and a total of N participants, with n_k participants in the k-th study, $k = 1, ..., K$, $\sum_{k=1}^{K} n_k = N$. Let D, G, E and S denote the

phenotype, genotype, environmental exposure and stratification factor. Let subscript (i,k) stand for participant i in study k, for $i = 1, ..., n_k; k = 1, ..., K$. Consider the following factorization of the retrospective likelihood $L^R = \prod_{k=1}^{K} L_k^R$,

$$\begin{aligned} L_k^R &= \prod_{i=1}^{n_k} P(G_{ki}, E_{ki}, S_{ki}|D_{ki}) \\ &= \prod_{i=1}^{n_k} \frac{P(D_{ki}|G_{ki}, E_{ki}, S_{ki})P(G_{ki}|E_{ki}, S_{ki})P(E_{ki}, S_{ki})}{\sum_{G_{ki}, E_{ki}, \boldsymbol{S}_{ki}} P(D_{ki}|G_{ki}, E_{ki}, S_{ki})P(G_{ki}|E_{ki}, S_{ki})P(E_{ki}, S_{ki})}. \end{aligned}$$

The three components $P(D|G,E,S)$, $P(G|E,S)$ and $P(E,S)$ in L_k^R are modeled in the following way:

(i) $P(D|G,E,S)$: We consider a logistic disease incidence model $P(D_{ki}|G_{ki}, E_{ki}, S_{ki}) = H\{\gamma_{0k} + \gamma_G G_{ki} + \gamma_E E_{ki} + \gamma_{GE} G_{ki} E_{ki} + \gamma_S S_{ki}\}$. Without loss of generality, we have assumed $\boldsymbol{\gamma} = (\gamma_G, \gamma_E, \gamma_{GE}, \gamma_S)$ is the common set of parameters of interest.

(ii) $P(G|E,S)$: For a dominant susceptibility model of G, we consider a logistic model

$$P(G_{ki} = 1|E_{ki}, S_{ki}) = H\{\eta_{0k} + \eta_k S_{ki} + \theta_k E_{ki}\}, \tag{5.1}$$

where θ_k (η_k) is a measure of dependence between G and E (S) in the k-th study. Under G-E independence conditional on S within each study k, model (5.1) can be reduced to

$$P(G_{ki} = 1|E_{ki}, S_{ki}) = H\{\eta_{0k} + \eta_k S_{ki}\}. \tag{5.2}$$

For additive susceptibility model of G, one may consider a proportional odds model to handle $P(G|E,S)$; similarly, for co-dominant susceptibility model, one may consider polychotomous logistic model. Alternatively, in this chapter, we consider a logistic model to handle the minor allele frequency (MAF) under the Hardy-Weinberg equilibrium (HWE) assumption (Chen et al. 2009). Let $q(E_{ki}, S_{ki})$ be the MAF for given (E_{ki}, S_{ki}). We model the MAF as $q(E_{ki}, S_{ki}) = H\{\eta_{0k} + \eta_k S_{ki} + \theta_k E_{ki}\}$, which can be reduced to $q(E_{ki}, S_{ki}) = H\{\eta_{0k} + \eta_k S_{ki}\}$ under G-E independence conditional on S_{ki}.

(iii) $P(E, S)$: The joint distribution function for (E, S) is allowed to remain completely nonparametric as in CC.

Remark 5.1. Let $\boldsymbol{\gamma}_0 = (\gamma_{01}, ..., \gamma_{0K})$, $(\boldsymbol{\eta}_0; \boldsymbol{\eta}) = (\eta_{01}, ..., \eta_{0K}; \eta_1, ..., \eta_K)$ and $\boldsymbol{\theta} = (\theta_1, ..., \theta_K)$. Let $\boldsymbol{\beta} = (\boldsymbol{\gamma}, \boldsymbol{\gamma}_0, \boldsymbol{\eta}, \boldsymbol{\eta}_0)$ and $\boldsymbol{\theta}$ be the full set of parameters under the unconstrained model. In models (i)-(iii), we consider $\boldsymbol{\gamma}$ as our common set of parameters across studies, $(\boldsymbol{\gamma}_0, \boldsymbol{\eta}, \boldsymbol{\eta}_0)$ as study level fixed nuisance parameters, and $\boldsymbol{\theta}$ as study level random nuisance parameters (under the EB framework). This formulation is flexible. For example, one may use a subvector of $\boldsymbol{\gamma}$ as the common parameters with the remaining of $\boldsymbol{\gamma}$ varying on k; one may also assume $\eta_k = \eta$ for $k = 1, ..., K$ in models (5.1) and (5.2), i.e., a common G-S association across studies, and treat $(\boldsymbol{\gamma}, \eta)$ as the common set of parameters. Then the methods in the following sections can be modified accordingly. Without loss of generality, we consider $\boldsymbol{\gamma}$ as our common set of disease odds ratio parameters that is of interest throughout this chapter.

Estimation. The MLE of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ (only $\boldsymbol{\beta}$ under the constrained model) can be obtained by the profile-likelihood techniques in CC. In particular, the MLEs under L^R can be equivalently obtained by maximizing a pseudo-likelihood L^*, in which estimation of the high dimensional nuisance parameters involved in the specification of $P(E, S)$ is not required. More details of the pseudo-likelihood method were provided in CC. The MLEs of $(\boldsymbol{\beta}, \boldsymbol{\theta})$ can be estimated using the R 'CGEN' package (Bhattacharjee et al. 2012).

5.2.2 Empirical Bayes estimator under multiple-study setting

In this section, we consider both IPD analysis and meta-analysis to handle the multiple-study problem. An IPD analysis can be performed when the individual level data are available. The MLEs of $\boldsymbol{\gamma}$ under the unconstrained and constrained models can be obtained directly by using the profile-likelihood techniques in CC, and are denoted as $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}^0$

respectively. We then propose an EB shrinkage estimator of $\boldsymbol{\gamma}$ that combines $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}^0$ for an IPD analysis. When the IPD are not available, e.g. in published results, we alternatively consider meta-analysis that combines study level summary statistics across studies. Based on the data from each study k, one collects the MLE $\tilde{\boldsymbol{\gamma}}_k$ (with covariance matrix $\tilde{\boldsymbol{V}}_{\tilde{\gamma}_k}$) and $\tilde{\theta}_k$ (with variance $\tilde{\sigma}^2_{\tilde{\theta}_k}$) from the unconstrained model; and collects $\tilde{\boldsymbol{\gamma}}^0_k$ and $\tilde{\boldsymbol{V}}_{\tilde{\gamma}^0_k}$ from the constrained model. We then consider EB shrinkage estimators that combine these study level summary statistics under a meta-analysis framework, where the shrinkage factors are estimated by borrowing strength across studies.

Under both IPD analysis and meta-analysis framework, we propose two variants of the EB estimator, MC type and partially Bayes (PB) type, and show consistency between the two approaches under certain formulations/assumptions. Throughout this chapter, for the EB estimator, we consider a strategy for conducting inference on $\boldsymbol{\gamma}$ with prior distribution on $\boldsymbol{\theta}$ and no further prior specification on $\boldsymbol{\beta}$. We consider $\theta_k \overset{iid}{\sim} N(0, \tau^2)$, in matrix notation, i.e., $\boldsymbol{\theta} \sim MVN(\mathbf{0}, \boldsymbol{A})$ where $\boldsymbol{A} = diag(\tau^2, ..., \tau^2)_{(K\times K)}$. Here $diag(a_1, ..., a_K)_{(K\times K)}$ stand for a $K \times K$ diagonal matrix whose diagonal entries are $(a_1, ..., a_K)$. In order to borrow strength across studies, we have assumed the uncertainty around the G-E independence assumption is constant (τ^2) across studies.

IPD analysis

(i) MC-type EB estimator: We propose an EB shrinkage estimator, in the spirit of MC, combining $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}^0$ under the multiple-study setting. We worked with $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ directly rather than its limiting value $\boldsymbol{\beta}(\boldsymbol{\theta})$ as suggested by Meng (2010). Let $\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ be the subvector of $\hat{\boldsymbol{\beta}}(\boldsymbol{\theta})$ corresponding to $\boldsymbol{\gamma}$. Following Taylor's expansion, we have the linear approximation $\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta}) \sim MVN(\hat{\boldsymbol{\gamma}}(\mathbf{0}), \Delta^\top \boldsymbol{A} \Delta)$, where $\Delta^\top = \partial \hat{\boldsymbol{\gamma}}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\mathbf{0}}$. Following the invariant property of MLE, we have $\hat{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}}) \sim MVN(\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta}), \boldsymbol{V}_{\hat{\gamma}(\hat{\theta})})$. Under the Gaussian-

Gaussian model, we propose a Bayes estimator (posterior mean) of $\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ of the form

$$\Delta^\top \boldsymbol{A}\Delta\{\boldsymbol{V}_{\hat{\gamma}(\hat{\theta})} + \Delta^\top \boldsymbol{A}\Delta\}^{-1}\hat{\boldsymbol{\gamma}} + \boldsymbol{V}_{\hat{\gamma}(\hat{\theta})}\{\boldsymbol{V}_{\hat{\gamma}(\hat{\theta})} + \Delta^\top \boldsymbol{A}\Delta\}^{-1}\hat{\boldsymbol{\gamma}}^0, \tag{5.3}$$

where $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\gamma}}^0$ are the MLEs of $\boldsymbol{\gamma}$ under the unconstrained and constrained models respectively. As in MC, we consider an ad hoc EB estimator $\hat{\boldsymbol{\gamma}}_{\text{MC-EB}}$ for $\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ by plugging the estimates of the weights $\Delta^\top \boldsymbol{A}\Delta\{\boldsymbol{V}_{\hat{\gamma}(\hat{\theta})} + \Delta^\top \boldsymbol{A}\Delta\}^{-1}$ and $\boldsymbol{V}_{\hat{\gamma}(\hat{\theta})}\{\boldsymbol{V}_{\hat{\gamma}(\hat{\theta})} + \Delta^\top \boldsymbol{A}\Delta\}^{-1}$ in (5.3). The components of the weights can be obtained as follows:

(i.a) $\boldsymbol{V}_{\hat{\gamma}}$: We use $\hat{\boldsymbol{V}}_{\hat{\gamma}}$ for $\boldsymbol{V}_{\hat{\gamma}}$ in (5.3). Denote $\boldsymbol{I}$ as the full observed information matrix under the unconstrained model; denote $\hat{\boldsymbol{I}}$ as $\boldsymbol{I}$ evaluated at the MLE. $\hat{\boldsymbol{V}}_{\hat{\gamma}}$ can be obtained as the corresponding sub-matrices of $\hat{\boldsymbol{I}}^{-1}$.

(i.b) Δ: MC proposed to use an approximation $\boldsymbol{I}_{\theta\gamma}\boldsymbol{I}_{\gamma\gamma}^{-1}|_{\boldsymbol{\theta}=\boldsymbol{0}}$, where $\boldsymbol{I}_{\theta\gamma}$ and $\boldsymbol{I}_{\gamma\gamma}$ are the corresponding sub-matrices of $\boldsymbol{I}$. The appropriateness of using $\boldsymbol{I}_{\theta\gamma}\boldsymbol{I}_{\gamma\gamma}^{-1}|_{\boldsymbol{\theta}=\boldsymbol{0}}$ is verified in Appendix B.4. Note that, applying Taylor's expansion, we have $\hat{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}}) \approx \hat{\boldsymbol{\gamma}}^0 + \{\partial\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}|_{\boldsymbol{\theta}=\boldsymbol{0}}\hat{\boldsymbol{\theta}}$, and then we have an approximation $\{\partial\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}|_{\boldsymbol{\theta}=\boldsymbol{0}} \approx (\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}^0)\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}$. We use $\hat{\Delta}^\top = (\hat{\boldsymbol{\gamma}} - \hat{\boldsymbol{\gamma}}^0)\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}$ for Δ in (5.3). Note that $\hat{\Delta}$ depends only on the MLEs $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\gamma}}^0$ and $\hat{\boldsymbol{\theta}}$, which bypasses the calculation of $\boldsymbol{I}_{\theta\gamma}\boldsymbol{I}_{\gamma\gamma}^{-1}|_{\boldsymbol{\theta}=\boldsymbol{0}}$.

(i.c) $\boldsymbol{A}$: We use EB estimator $\hat{\boldsymbol{A}}$ for $\boldsymbol{A}$. Let $\hat{\theta}_k$ be the MLE of θ_k under the unconstrained model, with variance $\hat{\sigma}^2_{\theta_k}$. We assume $\hat{\theta}_k \mid \theta_k \sim N(\theta_k, \hat{\sigma}^2_{\theta_k})$. Note that marginally $\hat{\theta}_k \sim N(0, \tau^2 + \hat{\sigma}^2_{\theta_k})$. We consider EB estimator of $\boldsymbol{A}$ using the following strategies: (1) a conservative estimator $\hat{\boldsymbol{A}} = \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top$ as in MC; (2) $\hat{\boldsymbol{A}} = diag(\hat{\tau}^2, ..., \hat{\tau}^2)_{(K\times K)}$, where $\hat{\tau}^2$ is the MLE that maximizes the marginal likelihood of $\hat{\boldsymbol{\theta}}$; (3) $\hat{\boldsymbol{A}} = diag(\bar{\tau}^2, ..., \bar{\tau}^2)_{(K\times K)}$, where $\bar{\tau}^2$ is the estimated posterior mean of $\tau^2|\hat{\boldsymbol{\theta}}$ in the spirit of Morris (1983). The details are shown in Appendix B.4.

Following (i.a)-(i.c), the resulting MC-type EB estimator is given by

$$\hat{\boldsymbol{\gamma}}_{\text{MC-EB}} = \hat{\Delta}^\top \hat{\boldsymbol{A}}\hat{\Delta}\{\hat{\boldsymbol{V}}_{\hat{\gamma}} + \hat{\Delta}^\top \hat{\boldsymbol{A}}\hat{\Delta}\}^{-1}\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{V}}_{\hat{\gamma}}\{\hat{\boldsymbol{V}}_{\hat{\gamma}} + \hat{\Delta}^\top \hat{\boldsymbol{A}}\hat{\Delta}\}^{-1}\hat{\boldsymbol{\gamma}}^0. \tag{5.4}$$

Denote $\hat{\gamma}_{\text{MC-EB1}}$, $\hat{\gamma}_{\text{MC-EB2}}$ and $\hat{\gamma}_{\text{MC-EB3}}$ as the MC-type EB estimators of $\hat{\gamma}(\boldsymbol{\theta})$ with $\hat{\boldsymbol{A}}$ following strategies (1), (2) and (3) in (i.c) respectively.

(ii) Partially Empirical Bayes (PEB) estimator: An alternative strategy is to consider an EB estimator under the 'partially Bayes' framework of Meng (2010). We extend it into a more general case with multiple parameters and unrestricted variance-covariance matrix structure. In particular, under the unconstrained model, we have that

$$(5.5)\quad \begin{pmatrix} \hat{\gamma} \\ \hat{\boldsymbol{\theta}} \end{pmatrix} = MVN\left(\begin{bmatrix} \gamma \\ \boldsymbol{\theta} \end{bmatrix}, \begin{bmatrix} \boldsymbol{I}^{\gamma\gamma} & \boldsymbol{I}^{\gamma\theta} \\ \boldsymbol{I}^{\theta\gamma} & \boldsymbol{I}^{\theta\theta} \end{bmatrix} \right), \quad \text{where} \quad \begin{bmatrix} \boldsymbol{I}^{\gamma\gamma} & \boldsymbol{I}^{\gamma\theta} \\ \boldsymbol{I}^{\theta\gamma} & \boldsymbol{I}^{\theta\theta} \end{bmatrix}$$

is the model based asymptotic covariance matrix of $(\hat{\gamma}, \hat{\boldsymbol{\theta}})$ obtained from the corresponding blocks of $\boldsymbol{I}^{-1}$. Note that $\hat{\boldsymbol{I}}^{\gamma\gamma} = \hat{\boldsymbol{V}}_{\hat{\gamma}}$ and $\hat{\boldsymbol{I}}^{\theta\theta} = \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\theta}}}$. Before any prior knowledge, the profile MLE of $\boldsymbol{\gamma}$ for fixed $\boldsymbol{\theta}$ under model (5.5) can be derived as

$$(5.6)\qquad \hat{\gamma}(\boldsymbol{\theta}) = \hat{\gamma} + \boldsymbol{I}^{\gamma\theta}(\boldsymbol{I}^{\theta\theta})^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Given the data ($\hat{\gamma}$ and $\hat{\boldsymbol{\theta}}$), the only unknown quantity in $\hat{\gamma}(\boldsymbol{\theta})$ in (5.6) is $\boldsymbol{\theta}$. Under the PB framework, we can make inference on $\hat{\gamma}(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$. Let $\Gamma^\top = \boldsymbol{I}^{\gamma\theta}(\boldsymbol{I}^{\theta\theta})^{-1}|_{\boldsymbol{\theta}=\mathbf{0}}$. Substituting $\Gamma^\top$ for $\boldsymbol{I}^{\gamma\theta}(\boldsymbol{I}^{\theta\theta})^{-1}$ in (5.6), we have an approximated distribution $\hat{\gamma}(\boldsymbol{\theta}) \sim MVN(\hat{\gamma}(\mathbf{0}), \Gamma^\top \boldsymbol{A}\Gamma)$. We use the posterior expectation of $\hat{\gamma}(\boldsymbol{\theta})$ as the PB estimator of $\hat{\gamma}(\boldsymbol{\theta})$ (for a fixed $\boldsymbol{A}$), which can be given as

$$(5.7)\qquad \Gamma^\top \boldsymbol{A}\Gamma\{\boldsymbol{V}_{\hat{\gamma}} + \Gamma^\top \boldsymbol{A}\Gamma\}^{-1}\hat{\gamma} + \boldsymbol{V}_{\hat{\gamma}}\{\boldsymbol{V}_{\hat{\gamma}} + \Gamma^\top \boldsymbol{A}\Gamma\}^{-1}\hat{\gamma}^0.$$

Again, we consider an ad hoc EB estimator $\hat{\gamma}_{\text{PEB}}$ for $\hat{\gamma}(\boldsymbol{\theta})$ by plugging the estimated weights in (5.7). We use $\hat{\boldsymbol{V}}_{\hat{\gamma}}$ for $\boldsymbol{V}_{\hat{\gamma}}$. Following (5.6), $\hat{\gamma} = \hat{\gamma}^0 + \boldsymbol{I}^{\gamma\theta}(\boldsymbol{I}^{\theta\theta})^{-1}|_{\boldsymbol{\theta}=\mathbf{0}}\hat{\boldsymbol{\theta}}$, and then we can use $(\hat{\gamma} - \hat{\gamma}^0)\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}$ as an estimator of $\Gamma^\top$ in (5.7), i.e., $\hat{\Gamma}^\top = (\hat{\gamma} - \hat{\gamma}^0)\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}$. We use the EB estimator $\hat{\boldsymbol{A}}$ for $\boldsymbol{A}$ in (5.7) following the strategies described in section 5.2.2(i.c). The resulting PEB estimator $\hat{\gamma}_{\text{PEB}}$ can be written as

$$(5.8)\qquad \hat{\gamma}_{\text{PEB}} = \hat{\Gamma}^\top \hat{\boldsymbol{A}}\hat{\Gamma}\{\hat{\boldsymbol{V}}_{\hat{\gamma}} + \hat{\Gamma}^\top \hat{\boldsymbol{A}}\hat{\Gamma}\}^{-1}\hat{\gamma} + \hat{\boldsymbol{V}}_{\hat{\gamma}}\{\hat{\boldsymbol{V}}_{\hat{\gamma}} + \hat{\Gamma}^\top \hat{\boldsymbol{A}}\hat{\Gamma}\}^{-1}\hat{\gamma}^0,$$

Note that, since $\hat{\Gamma} = \hat{\Delta}$, we have $\hat{\gamma}_{\text{PEB}} = \hat{\gamma}_{\text{MC-EB}}$.

Meta-analysis

(i) MC-type EB estimator. We propose to use (multivariate) inverse variance weighted (IVW) estimator $\tilde{\gamma} = \{\sum_k \tilde{\boldsymbol{V}}_{\tilde{\gamma}_k}^{-1}\}^{-1} \sum_k \tilde{\boldsymbol{V}}_{\tilde{\gamma}_k}^{-1} \tilde{\gamma}_k$ to combine $\tilde{\gamma}_k$ under the unconstrained model across studies; and similarly use $\tilde{\gamma}^0 = \{\sum_k \tilde{\boldsymbol{V}}_{\tilde{\gamma}_k^0}^{-1}\}^{-1} \sum_k \tilde{\boldsymbol{V}}_{\tilde{\gamma}_k^0}^{-1} \tilde{\gamma}_k^0$ to combine $\tilde{\gamma}_k^0$ under the constrained model. In order to construct an EB estimator combining $\tilde{\gamma}$ and $\tilde{\gamma}^0$ in the spirit of MC, we first prove the following Lemma.

Lemma 1. For the unconstrained model, $\boldsymbol{V}_{\tilde{\gamma}} = \boldsymbol{V}_{\hat{\gamma}}$; $\tilde{\gamma}$ and $\hat{\gamma}$ have the same limiting normal distribution $MVN(\boldsymbol{\gamma}, \boldsymbol{V}_{\hat{\gamma}})$. For the constrained model, $\boldsymbol{V}_{\tilde{\gamma}^0} = \boldsymbol{V}_{\hat{\gamma}^0}$; $\tilde{\gamma}^0$ and $\hat{\gamma}^0$ have the same limiting normal distribution $MVN(\boldsymbol{\gamma}(\mathbf{0}), \boldsymbol{V}_{\hat{\gamma}^0})$.

Proof: Follow Lin and Zeng (2010).

Lemma 2. $\tilde{\theta}_k = \hat{\theta}_k$ and $\tilde{\sigma}_{\theta_k}^2 \geq \hat{\sigma}_{\theta_k}^2$, for $k = 1, ..., K$.

Proof: shown in Appendix B.4.

Because $\hat{\gamma} = \hat{\gamma}(\hat{\boldsymbol{\theta}}) \sim MVN(\hat{\gamma}(\boldsymbol{\theta}), \boldsymbol{V}_{\hat{\gamma}(\hat{\theta})})$, along with Lemma 1, we have $\tilde{\gamma} \sim MVN(\hat{\gamma}(\boldsymbol{\theta}), \boldsymbol{V}_{\tilde{\gamma}})$. Under prior distribution $\boldsymbol{\theta} \sim MVN(\mathbf{0}, \boldsymbol{A})$, we can obtain an approximated distribution for $\hat{\gamma}(\boldsymbol{\theta})$ as $MVN(\hat{\gamma}(\mathbf{0}), \Delta^\top \boldsymbol{A} \Delta)$ following Taylor's expansion. Along with Lemma 1, $\hat{\gamma}(\boldsymbol{\theta}) \sim MVN(\tilde{\gamma}^0, \Delta^\top \boldsymbol{A} \Delta)$ asymptotically. For the Gaussian-Gaussian model, we consider an MC-type EB estimator combining $\tilde{\gamma}$ and $\tilde{\gamma}^0$ as

$$\tilde{\gamma}_{\text{MC-EB}} = \tilde{\Delta}^\top \tilde{\boldsymbol{A}} \tilde{\Delta} \{\tilde{\boldsymbol{V}}_{\tilde{\gamma}} + \tilde{\Delta}^\top \tilde{\boldsymbol{A}} \tilde{\Delta}\}^{-1} \tilde{\gamma} + \tilde{\boldsymbol{V}}_{\tilde{\gamma}} \{\tilde{\boldsymbol{V}}_{\tilde{\gamma}} + \tilde{\Delta}^\top \tilde{\boldsymbol{A}} \tilde{\Delta}\}^{-1} \tilde{\gamma}^0. \tag{5.9}$$

Again, estimated weights have been used in (5.9). In particular, $\tilde{\boldsymbol{V}}_{\tilde{\gamma}} = \{\sum_k \tilde{\boldsymbol{V}}_{\tilde{\gamma}_k}^{-1}\}^{-1}$, $\tilde{\boldsymbol{A}}$ is the EB estimator of $\boldsymbol{A}$ calculated based on the study level statistics $\tilde{\theta}_k$ and $\tilde{\sigma}_{\theta_k}^2$ under the strategies described in section 5.2.2(i.c), and $\tilde{\Delta} = (\tilde{\gamma} - \tilde{\gamma}^0)\tilde{\boldsymbol{\theta}}^\top (\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^\top)^{-1}$. Following Lemma 1 and 2, $\tilde{\Delta}$ and $\hat{\Delta}$ have the same limiting value, so we have used $\tilde{\Delta}$ as an approximate

estimator for Δ. Note that all the components in $\tilde{\gamma}_{\text{MC-EB}}$ in (5.9) are obtained using only study level summary statistics.

(ii) PEB estimator. Note that the k-th study only involves parameters $\boldsymbol{\beta}_k = (\boldsymbol{\gamma}, \gamma_{0k}, \eta_{0k}, \eta_k)$ and θ_k. Let $\tilde{\boldsymbol{\beta}}_k(\theta_k)$ be the profile MLE of $\boldsymbol{\beta}_k$ for fixed θ_k for the k-study, and $\tilde{\boldsymbol{\gamma}}_k(\theta_k)$ be the sub-vector of $\tilde{\boldsymbol{\beta}}_k(\theta_k)$ corresponding to $\boldsymbol{\gamma}$. From each study k, we have the study level data $\tilde{\boldsymbol{\gamma}}_k$, $\tilde{\theta}_k$ and the full observed information matrix $\tilde{\boldsymbol{I}}_k$ under the unconstrained model, and $\tilde{\boldsymbol{\gamma}}_k^0$ under the constrained model, for $k = 1, ..., K$. Under the unconstrained model, we have that

$$\begin{pmatrix} \tilde{\boldsymbol{\gamma}}_k \\ \tilde{\theta}_k \end{pmatrix} = MVN\left(\begin{bmatrix} \boldsymbol{\gamma} \\ \theta_k \end{bmatrix}, \begin{bmatrix} \boldsymbol{I}^{k\gamma\gamma} & \boldsymbol{I}^{k\gamma\theta_k} \\ \boldsymbol{I}^{k\theta_k\gamma} & \boldsymbol{I}^{k\theta_k\theta_k} \end{bmatrix} \right), \quad \text{where} \quad \begin{bmatrix} \boldsymbol{I}^{k\gamma\gamma} & \boldsymbol{I}^{k\gamma\theta_k} \\ \boldsymbol{I}^{k\theta_k\gamma} & \boldsymbol{I}^{k\theta_k\theta_k} \end{bmatrix} \tag{5.10}$$

is the model based asymptotic covariance matrix of $(\tilde{\boldsymbol{\gamma}}_k, \tilde{\theta}_k)$ obtained from the corresponding blocks of $\boldsymbol{I}_k^{-1}$. Note that $\tilde{\boldsymbol{I}}^{k\gamma\gamma} = \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{\gamma}}_k}$ and $\tilde{\boldsymbol{I}}^{k\theta_k\theta_k} = \tilde{\sigma}^2_{\theta_k}$. Before any prior knowledge, the MLE of $\boldsymbol{\gamma}$ under model (5.10) is $\{\sum_k \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{\gamma}}_k}^{-1}\}^{-1} \sum_k \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{\gamma}}_k}^{-1} \tilde{\boldsymbol{\gamma}}_k$, i.e. $\tilde{\boldsymbol{\gamma}}$; the MLE of θ_k is $\tilde{\theta}_k$. The conditional distribution $\tilde{\boldsymbol{\gamma}}_k | \tilde{\theta}_k \sim MVN(\boldsymbol{\gamma} + \boldsymbol{I}^{k\gamma\theta_k}(\boldsymbol{I}^{k\theta_k\theta_k})^{-1}(\tilde{\theta}_k - \theta_k), \boldsymbol{I}^{k\gamma\gamma} - \boldsymbol{I}^{k\gamma\theta_k}(\boldsymbol{I}^{k\theta_k\theta_k})^{-1}\boldsymbol{I}^{k\theta_k\gamma})$. Then we have the profile MLE $\tilde{\boldsymbol{\gamma}}_k(\theta_k) = \tilde{\boldsymbol{\gamma}}_k + \boldsymbol{I}^{k\gamma\theta_k}(\boldsymbol{I}^{k\theta_k\theta_k})^{-1}(\theta_k - \tilde{\theta}_k)$, with variance $\boldsymbol{V}_{\tilde{\boldsymbol{\gamma}}_k(\theta_k)} = \boldsymbol{I}^{k\gamma\gamma} - \boldsymbol{I}^{k\gamma\theta_k}(\boldsymbol{I}^{k\theta_k\theta_k})^{-1}\boldsymbol{I}^{k\theta_k\gamma}$ for fixed θ_k and $\boldsymbol{V}_{\tilde{\boldsymbol{\gamma}}_k(\tilde{\theta}_k)} = \boldsymbol{I}^{k\gamma\gamma}$ for $\theta_k = \tilde{\theta}_k$. It follows that the profile MLE $\tilde{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ under model (5.10) can be derived as

$$\tilde{\boldsymbol{\gamma}}(\boldsymbol{\theta}) = \Big\{ \sum_k \boldsymbol{V}_{\tilde{\boldsymbol{\gamma}}_k(\theta_k)}^{-1} \Big\}^{-1} \sum_k \boldsymbol{V}_{\tilde{\boldsymbol{\gamma}}_k(\theta_k)}^{-1} \tilde{\boldsymbol{\gamma}}_k(\theta_k). \tag{5.11}$$

Then $\tilde{\boldsymbol{\gamma}}(\tilde{\boldsymbol{\theta}}) = \{\sum_k \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{\gamma}}_k(\tilde{\theta}_k)}^{-1}\}^{-1} \sum_k \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{\gamma}}_k(\tilde{\theta}_k)}^{-1} \tilde{\boldsymbol{\gamma}}_k(\tilde{\theta}_k) = \tilde{\boldsymbol{\gamma}}$ and $\tilde{\boldsymbol{\gamma}}(\mathbf{0}) = \{\sum_k \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{\gamma}}_k^0}^{-1}\}^{-1} \sum_k \tilde{\boldsymbol{V}}_{\tilde{\boldsymbol{\gamma}}_k^0}^{-1} \tilde{\boldsymbol{\gamma}}_k^0 = \tilde{\boldsymbol{\gamma}}^0$.

Given the study level data $\tilde{\boldsymbol{\gamma}}_k$ and $\tilde{\theta}_k$, the only unknown quantity in $\tilde{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ in (5.11) is $\boldsymbol{\theta}$. Under the PB framework, we can make inference on $\tilde{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$. Again, applying Taylor's expansion at $\boldsymbol{\theta} = \mathbf{0}$, we have the linear approximation $\tilde{\boldsymbol{\gamma}}(\boldsymbol{\theta}) \approx \tilde{\boldsymbol{\gamma}}(\mathbf{0}) +$

$\{\partial\tilde{\gamma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}|_{\boldsymbol{\theta}=\mathbf{0}}\boldsymbol{\theta}$, and then we have the approximated distribution $\tilde{\gamma}(\boldsymbol{\theta}) \sim MVN(\tilde{\gamma}(\mathbf{0}),$ $\Lambda^\top \boldsymbol{A}\Lambda)$, where $\Lambda^\top = \partial\tilde{\gamma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\mathbf{0}}$. Since $\tilde{\gamma} \approx \tilde{\gamma}^0 + \{\partial\tilde{\gamma}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\}|_{\boldsymbol{\theta}=\mathbf{0}}\tilde{\boldsymbol{\theta}}$, we can approximate $\Lambda^\top$ by $(\tilde{\gamma} - \tilde{\gamma}^0)\tilde{\boldsymbol{\theta}}^\top(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^\top)^{-1}$ to avoid the calculation of the derivative of the profile function $\tilde{\gamma}(\boldsymbol{\theta})$. We consider a PEB estimator $\tilde{\gamma}_{\text{PEB}}$ for $\tilde{\gamma}(\boldsymbol{\theta})$ as

$$\tilde{\gamma}_{\text{PEB}} = \tilde{\Lambda}^\top\tilde{\boldsymbol{A}}\tilde{\Lambda}\{\tilde{\boldsymbol{V}}_{\tilde{\gamma}} + \tilde{\Lambda}^\top\tilde{\boldsymbol{A}}\tilde{\Lambda}\}^{-1}\tilde{\gamma} + \tilde{\boldsymbol{V}}_{\tilde{\gamma}}\{\tilde{\boldsymbol{V}}_{\tilde{\gamma}} + \tilde{\Lambda}^\top\tilde{\boldsymbol{A}}\tilde{\Lambda}\}^{-1}\tilde{\gamma}^0, \tag{5.12}$$

Again, estimated weights have been used in (5.9), where $\tilde{\boldsymbol{V}}_{\tilde{\gamma}} = \left\{\sum_k \tilde{\boldsymbol{V}}_{\tilde{\gamma}_k}^{-1}\right\}^{-1}$, $\tilde{\Lambda}^\top = (\tilde{\gamma} - \tilde{\gamma}^0)\tilde{\boldsymbol{\theta}}^\top(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^\top)^{-1}$, and $\tilde{\boldsymbol{A}}$ is the EB estimator of $\boldsymbol{A}$ calculated based on the study level statistics $\tilde{\theta}_k$ and $\tilde{\sigma}^2_{\tilde{\theta}_k}$ under the strategies described in section 5.2.2(i.c). Since $\tilde{\Lambda} = \tilde{\Delta}$, we have $\tilde{\gamma}_{\text{PEB}} = \tilde{\gamma}_{\text{MC-EB}}$.

Remark 5.2: *Equivalence between MC-type EB and PEB estimator.* Among the proposed EB estimators, we note that the MC-type EB and PEB estimator are equivalent under either IPD analysis or meta-analysis, i.e., $\hat{\gamma}_{\text{MC-EB}} = \hat{\gamma}_{\text{PEB}}$ and $\tilde{\gamma}_{\text{MC-EB}} = \tilde{\gamma}_{\text{PEB}}$. Therefore, we no longer distinguish between MC-type EB and PEB estimators under the Gaussian-Gaussian model. Denote the proposed EB estimator under the IPD analysis and meta-analysis as $\hat{\gamma}_{\text{EB}}$ and $\tilde{\gamma}_{\text{EB}}$. Under the IPD analysis (meta-analysis) framework, denote $\hat{\gamma}_{\text{EB1}}$, $\hat{\gamma}_{\text{EB2}}$ and $\hat{\gamma}_{\text{EB3}}$ ($\tilde{\gamma}_{\text{EB1}}$, $\tilde{\gamma}_{\text{EB2}}$ and $\tilde{\gamma}_{\text{EB3}}$) as the EB estimators with $\hat{\boldsymbol{A}}$ ($\tilde{\boldsymbol{A}}$) following strategies (1), (2) and (3) in (i.c) respectively.

Remark 5.3: *The conservative estimator $\hat{\boldsymbol{A}} = \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top$.* Let $\hat{\psi} = \hat{\gamma} - \hat{\gamma}^0$. For the conservative estimator $\hat{\boldsymbol{A}} = \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top$, then $\hat{\Delta}^\top\hat{\boldsymbol{A}}\hat{\Delta} = \hat{\psi}\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}\hat{\boldsymbol{A}}(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}\hat{\boldsymbol{\theta}}\hat{\psi}^\top = \hat{\psi}\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}\hat{\boldsymbol{\theta}}\hat{\psi}^\top$ $= \hat{\psi}\hat{\psi}^\top$. Here it is easy to verify that $\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}\hat{\boldsymbol{\theta}} = 1$. Note that $\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}\hat{\boldsymbol{\theta}}$ is a scalar. Let $\hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}\hat{\boldsymbol{\theta}} = \lambda$, then $\lambda\hat{\boldsymbol{\theta}}^\top = \hat{\boldsymbol{\theta}}^\top(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top)^{-1}\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top = \hat{\boldsymbol{\theta}}^\top$, so $\lambda = 1$. For the IPD analysis, we have a conservative EB estimator $\hat{\gamma}_{\text{EB1}}$ of the form

$$\hat{\gamma}_{\text{EB1}} = \hat{\psi}\hat{\psi}^\top\{\hat{\boldsymbol{V}}_{\hat{\gamma}} + \hat{\psi}\hat{\psi}^\top\}^{-1}\hat{\gamma} + \hat{\boldsymbol{V}}_{\hat{\gamma}}\{\hat{\boldsymbol{V}}_{\hat{\gamma}} + \hat{\psi}\hat{\psi}^\top\}^{-1}\hat{\gamma}^0. \tag{5.13}$$

Similarly, if we use the conservative estimator $\tilde{\boldsymbol{A}} = \tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^\top$, then $\tilde{\Delta}^\top\tilde{\boldsymbol{A}}\tilde{\Delta} = \tilde{\psi}\tilde{\boldsymbol{\theta}}^\top(\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^\top)^{-1}\tilde{\boldsymbol{\theta}}\tilde{\psi}^\top$

$= \tilde{\psi}\tilde{\psi}^\top$. For the meta-analysis, we have a conservative EB estimator $\tilde{\gamma}_{\text{EB1}}$ of the form

$$\tilde{\gamma}_{\text{EB1}} = \tilde{\psi}\tilde{\psi}^\top\{\tilde{V}_{\tilde{\gamma}} + \tilde{\psi}\tilde{\psi}^\top\}^{-1}\tilde{\gamma} + \tilde{V}_{\tilde{\gamma}}\{\tilde{V}_{\tilde{\gamma}} + \tilde{\psi}\tilde{\psi}^\top\}^{-1}\tilde{\gamma}^0. \tag{5.14}$$

According to lemma 1 and 2, $\tilde{\gamma}$ and $\hat{\gamma}$ have the same limiting normal distribution; $\tilde{\gamma}^0$ and $\hat{\gamma}^0$ have the same limiting normal distribution; the corresponding weights have the same limiting value. Thus, $\hat{\gamma}_{\text{EB1}}$ and $\tilde{\gamma}_{\text{EB1}}$ converge to the same limiting distribution.

(iii) Inverse variance weighted approach. From each study k, one can obtain an EB estimator $\tilde{\gamma}_{\text{EBk}}$ following MC as

$$\tilde{\gamma}_{\text{EBk}} = \tilde{\Delta}_k^\top \tilde{A}_k \tilde{\Delta}_k\{\tilde{V}_{\tilde{\gamma}_k} + \tilde{\Delta}_k^\top \tilde{A}_k \tilde{\Delta}_k\}^{-1}\tilde{\gamma}_k + \tilde{V}_{\tilde{\gamma}_k}\{\tilde{V}_{\tilde{\gamma}_k} + \tilde{\Delta}_k^\top \tilde{A}_k \tilde{\Delta}_k\}^{-1}\tilde{\gamma}_k^0,$$

where $\tilde{\Delta}_k^\top \tilde{A}_k \tilde{\Delta}_k$ is calculated similarly as before but using the data from study k only. In particular, $\tilde{\Delta}_k^\top = (\tilde{\gamma}_k - \tilde{\gamma}_k^0)\tilde{\theta}_k^{-1}$ and $\tilde{A}_k$ follows (i.c), e.g. $\tilde{A}_k = \tilde{\theta}_k^2$ for the conservative estimator. A naive approach to obtain a pooled EB estimator across studies is just to use the IVW average of $\tilde{\gamma}_{\text{EBk}}$, which is given by

$$\tilde{\gamma}_{\text{IVW-EB}} = \Big\{\sum_k \tilde{V}_{\tilde{\gamma}_{\text{EBk}}}^{-1}\Big\}^{-1} \sum_k \tilde{V}_{\tilde{\gamma}_{\text{EBk}}}^{-1}\tilde{\gamma}_{\text{EBk}}.$$

Alternatively, we can estimate the approximated prior variance $\Delta^\top A \Delta$ of $\hat{\gamma}(\boldsymbol{\theta})$ borrowing strength across all the study, and consider an hybrid EB estimator $\tilde{\gamma}_{\text{EBk}}$ of the form

$$\tilde{\gamma}'_{\text{EBk}} = \tilde{\Delta}^\top \tilde{A} \tilde{\Delta}\{\tilde{V}_{\tilde{\gamma}_k} + \tilde{\Delta}^\top \tilde{A} \tilde{\Delta}\}^{-1}\tilde{\gamma}_k + \tilde{V}_{\tilde{\gamma}_k}\{\tilde{V}_{\tilde{\gamma}_k} + \tilde{\Delta}^\top \tilde{A} \tilde{\Delta}\}^{-1}\tilde{\gamma}_k^0$$

for each study k. Here, $\tilde{\Delta}^\top \tilde{A} \tilde{\Delta}$ in $\tilde{\gamma}'_{\text{EBk}}$ is the same as that in (5.9). We consider hybrid IVW (HIVW) EB estimator given by

$$\tilde{\gamma}_{\text{HIVW-EB}} = \Big\{\sum_k \tilde{V}_{\tilde{\gamma}'_{\text{EBk}}}^{-1}\Big\}^{-1} \sum_k \tilde{V}_{\tilde{\gamma}'_{\text{EBk}}}^{-1}\tilde{\gamma}'_{\text{EBk}}.$$

5.2.3 Asymptotic variance

Under the IPD analysis, for $\hat{\gamma}_{\text{EB1}}$, the variance $\hat{V}_{\hat{\gamma}_{EB1}}$ can be obtained by viewing $\hat{\gamma}_{\text{EB1}}$ as a function of the MLE $(\hat{\gamma}, \hat{\boldsymbol{\theta}}, \hat{\gamma}^0)$. The joint asymptotic multivariate normal distri-

bution for these three estimates can be obtained in terms of the associated score functions and information matrices. An application of the Delta method provides the asymptotic variance-covariance matrix for $\hat{\gamma}_{\text{EB1}}$. For $\hat{\gamma}_{\text{EB2}}$ and $\hat{\gamma}_{\text{EB3}}$ of the more general form $\hat{\gamma}_{\text{EB}} = \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta} \{\hat{\boldsymbol{V}}_{\hat{\gamma}} + \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta}\}^{-1} \hat{\gamma} + \hat{\boldsymbol{V}}_{\hat{\gamma}} \{\hat{\boldsymbol{V}}_{\hat{\gamma}} + \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta}\}^{-1} \hat{\gamma}^0$, it is hard to apply the Delta method to obtain a variance formula for the EB estimator since $\hat{\boldsymbol{A}}$ does not have a closed form expression in terms of $(\hat{\gamma}, \hat{\boldsymbol{\theta}}, \hat{\gamma}^0)$. Instead, we consider an *ad hoc* way to calculate the variance. The derivation and expression of $\hat{\boldsymbol{V}}_{\hat{\gamma}_{EB1}}$, $\hat{\boldsymbol{V}}_{\hat{\gamma}_{EB2}}$ and $\hat{\boldsymbol{V}}_{\hat{\gamma}_{EB3}}$ is deferred to Appendix B.4. A Wald-type test can be constructed based on the statistic $\hat{\gamma}_{EB}^\top \hat{\boldsymbol{V}}_{\hat{\gamma}_{EB}}^{-1} \hat{\gamma}_{EB}$.

Under the meta-analysis scenario, $\tilde{\boldsymbol{V}}_{\tilde{\gamma}_{\text{EB1}}}$, $\tilde{\boldsymbol{V}}_{\tilde{\gamma}_{\text{EB2}}}$ and $\tilde{\boldsymbol{V}}_{\tilde{\gamma}_{\text{EB3}}}$ can be obtained similarly. For $\tilde{\gamma}_{\text{IVW-EB}}$, $\hat{\boldsymbol{V}}_{\tilde{\gamma}_{\text{IVW-EB}}} = \left\{ \sum_k \tilde{\boldsymbol{V}}_{\tilde{\gamma}_{\text{EBk}}}^{-1} \right\}^{-1}$. The validity of the above form of $\hat{\boldsymbol{V}}_{\tilde{\gamma}_{\text{IVW-EB}}}$ requires that the true asymptotic variance $\boldsymbol{V}_{\tilde{\gamma}_{\text{EBk}}}$ is estimated with negligible error. We note that this assumption is often implicitly assumed to hold in classic meta-analysis literature dealing with Gaussian models (e.g., DerSimonian and Laird (1986), Whitehead and Whitehead (1991), LZ). As the components of $\hat{\boldsymbol{V}}_{\tilde{\gamma}_{\text{IVW-EB}}}$, $\tilde{\boldsymbol{V}}_{\tilde{\gamma}_{\text{EBk}}}$ for the EB estimator $\tilde{\gamma}_{\text{EBk}}$ was derived using a Delta's method as an approximation, we would evaluate the performance of $\hat{\boldsymbol{V}}_{\tilde{\gamma}_{\text{IVW-EB}}}$ by comparing it with its empirical value, and similarly for $\hat{\boldsymbol{V}}_{\tilde{\gamma}_{\text{HIVW-EB}}}$.

5.2.4 Empirical linear Bayes rule

MC-type EB estimator is derived under the usual Gaussian-Gaussian hierarchical model, however, the PB approach can be generalized to non-normal cases. In particular, for a general prior on $\boldsymbol{\theta}$, we can derive the posterior of $\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}$, and then make inference on $\hat{\gamma}(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}$. We use the posterior expectation $E_\theta\{\hat{\gamma}(\boldsymbol{\theta})\} = \hat{\gamma} + E_\theta\{\boldsymbol{I}^{\gamma\theta}(\boldsymbol{I}^{\theta\theta})^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\}$ as the PB estimator of $\hat{\gamma}(\boldsymbol{\theta})$, where E_θ is the expectation taken with respect to the posterior distribution of $\boldsymbol{\theta}$ (for fixed hyperparameters of $\boldsymbol{\theta}$, e.g. τ^2). If we replace $\boldsymbol{I}^{\gamma\theta}(\boldsymbol{I}^{\theta\theta})^{-1}$ by $\hat{\boldsymbol{I}}^{\gamma\theta}(\hat{\boldsymbol{I}}^{\theta\theta})^{-1}$, then $E_\theta\{\hat{\gamma}(\boldsymbol{\theta})\} = \hat{\gamma} + \hat{\boldsymbol{I}}^{\gamma\theta}(\hat{\boldsymbol{I}}^{\theta\theta})^{-1}\{E_\theta(\boldsymbol{\theta}) - \hat{\boldsymbol{\theta}}\}$.

Consider the model $\hat{\theta}_k \mid \theta_k \sim N(\theta_k, \hat{\sigma}^2_{\theta_k})$ and $\theta_k \overset{iid}{\sim} (M_k, V_k)$. The notation indicating θ_k has mean M_k and variance V_k with no other assumptions about the distribution on θ_k. Let $W_k = V_k/(V_k + \hat{\sigma}^2_{\theta_k})$. Following Efron (1973), we have the 'linear Bayes rule' $M_k + W_k(\hat{\theta}_k - M_k)$ for the posterior mean of $\theta_k \mid \hat{\theta}_k$. Let $\boldsymbol{M} = (M_1, ..., M_K)^\top$, $\boldsymbol{V} = (V_1, ..., V_K)^\top$ and $\boldsymbol{W} = diag(W_1, ..., W_K)_{(K\times K)}$. The 'linear Bayes rule' for $\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}$ in matrix form is $\boldsymbol{M} + \boldsymbol{W}(\hat{\boldsymbol{\theta}} - \boldsymbol{M})$. Thus, the partially linear Bayes (PLB) estimator for fixed $\boldsymbol{M}$ and $\boldsymbol{V}$ can be given by

$$\hat{\boldsymbol{\gamma}}_{\text{PLB}} = \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{I}}^{\gamma\theta}(\hat{\boldsymbol{I}}^{\theta\theta})^{-1}(1 - \boldsymbol{W})(\boldsymbol{M} - \hat{\boldsymbol{\theta}}),$$

It is easy to show that $\boldsymbol{V}_{\hat{\gamma}_{PLB}} \leq \boldsymbol{V}_{\boldsymbol{\gamma}} - \boldsymbol{I}^{\gamma\theta}(\boldsymbol{I}^{\theta\theta})^{-1}(1 - \boldsymbol{W})\boldsymbol{I}^{\theta\theta}(1 - \boldsymbol{W})(\boldsymbol{I}^{\theta\theta})^{-1}\boldsymbol{I}^{\theta\gamma} \leq \boldsymbol{V}_{\hat{\boldsymbol{\gamma}}}$. Here we follow the definition that a matrix is greater than the other iff the corresponding subtracted matrix is symmetric positive definite (SPD). So $\hat{\boldsymbol{\gamma}}_{\text{PLB}}$ has more precision than $\hat{\boldsymbol{\gamma}}$. If $\boldsymbol{M}$ and $\boldsymbol{W}$ are unknown, we can use the 'empirical linear Bayes rule' $\hat{\boldsymbol{M}} + \hat{\boldsymbol{W}}(\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{M}})$ with $\hat{\boldsymbol{M}}$ and $\hat{\boldsymbol{W}}$ estimated from the data (Efron 1973). Then it results in a partially empirical linear Bayes (PELB) estimator of $\hat{\boldsymbol{\gamma}}(\boldsymbol{\theta})$ of the form

$$\hat{\boldsymbol{\gamma}}_{\text{PELB}} = \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{I}}^{\gamma\theta}(\hat{\boldsymbol{I}}^{\theta\theta})^{-1}\{(1 - \hat{\boldsymbol{W}})(\hat{\boldsymbol{M}} - \hat{\boldsymbol{\theta}})\}.$$

We consider a typical example of a mixture distribution $\theta_k \sim p\,\delta(0) + (1 - p)N(0, \tau^2)$, where $\delta(0)$ is a point mass at 0 reflecting G-E independence. The population level G-E association structure is assumed as a mixture distribution reflecting that a fraction, say p, of the studies have G-E independence holding, whereas the remaining studies (a fraction of $1 - p$) show some departures from the independence assumption. For this distribution, we have $M_k = 0$ and $V_k = (1 - p)\tau^2$ for $k = 1, ..., K$. So the 'empirical linear Bayes rule' is $\hat{\boldsymbol{W}}\hat{\boldsymbol{\theta}}$, where we use $\hat{W}_k = \frac{(1-\hat{p})\hat{\tau}^2}{(1-\hat{p})\hat{\tau}^2 + \hat{\sigma}^2_{\theta_k}}$ with $\hat{p}$ and $\hat{\tau}^2$ obtained by maximizing the marginal likelihood of $\hat{\theta}_k \sim pN(0, \hat{\sigma}^2_{\theta_k}) + (1 - p)N(0, \hat{\sigma}^2_{\theta_k} + \tau^2)$. The details are shown in Appendix B.4.

5.3 A Simulation Study

Simulation scenarios

We considered a simulation study to compare the relative performances of the proposed methods, including (a) the standard logistic regression analysis; (b) the constrained and unconstrained estimator based on the retrospective likelihood method; (c) the proposed EB estimator (EB1, EB2 and EB3) that combines the constrained and unconstrained estimator. The relative performances were also compared under both IPD and meta-analytic setting.

We considered $\boldsymbol{S} = (S_1, S_2)$, where S_1 is a binary variable with $P(S_1 = 1) = 0.55$ and $P(S_1 = 0) = 0.45$, and S_2 is a continuous variable follows $N(0, 1)$. We generated the environmental covariate as $E = min(10, \exp(X))$, where X follows a normal distribution with the mean parameter depends on S_1. In particular, $X|S_1 = 0 \sim N(0, 1)$ and $X|S_1 = 1 \sim N(0.05, 1)$. Conditional on $(E, \boldsymbol{S})$, we generated a binary genetic factor G through a logistic model of the form

$$P(G_{ki} = 1|E_{ki}, \boldsymbol{S}_{ki}) = H\{\eta^*_{0k} + \eta^{*\top}_k \boldsymbol{S}_{ki} + \theta^*_k E_{ki}\},$$

where $\eta^{*\top}_k = (\eta^*_{1k}, \eta^*_{2k})$. We set $\eta^*_{1k} \overset{iid}{\sim}$ Uniform$(0.1, 0.3)$ and $\eta^*_{2k} \overset{iid}{\sim}$ Uniform$(0, 0.2)$ to reflect a strong G-$\boldsymbol{S}$ association. To explore the role of departure from G-E independence (conditional on $\boldsymbol{S}$), the G-E association is set as $\theta^*_k = 0$ corresponding to G-E independence; and set as $\theta^*_k \overset{iid}{\sim}$ Uniform$(-c, c)$ corresponding to G-E dependence, where we considered $c = 0.2$ and 0.5 corresponding to modest and strong G-E associations. We also considered the situation where G-E associations follow a mixture distribution $\theta^*_k \sim p\,\delta(0) + (1 - p)$Uniform$(-c, c)$ (choices of p and c are given in Table 5.3). Here a fraction (say p) of the studies have G-E independence holding, whereas the remaining studies show some departures from the independence assumption. We set $\eta^*_{0k} \overset{iid}{\sim}$ Uniform$(-1.6, -1.3)$ so that the MAF of G is 0.2. Conditional on $(G, E, \boldsymbol{S})$, we

generated a binary disease outcome D from the logistic regression model

$$P(D_{ki}=1|G_{ki},E_{ki},\boldsymbol{S}_{ki})=H\{\gamma_{0k}^{*}+\gamma_{G}^{*}G_{ki}+\gamma_{E}^{*}E_{ki}+\gamma_{GE}^{*}G_{ki}E_{ki}+\boldsymbol{\gamma}_{S}^{*\top}\boldsymbol{S}_{ki}\}.$$

We set $(\gamma_G^*, \gamma_E^*, \gamma_{GE}^*; \boldsymbol{\gamma}_S^{*\top}) = (0.2, 0.1, 0.1; 0.1, 0.05)$ to reflect small to modest main and interaction effects between G and E, along with a modest stratification risk factor $\boldsymbol{S}$. We set $\gamma_{0k}^* \overset{iid}{\sim} \text{Uniform}(-4.6, -4.2)$ such that the marginal probability of the disease in the population is 2%. When $\theta_k^* = 0$, the low disease prevalence leads to G-E independence in controls approximately.

We generated data $(D, G, E, \boldsymbol{S})$ from a large cohort including a total of $5,000,000$ patients, divided by $K = 10$ sub-cohorts with different sample sizes ($N_1 = 1,000,000$; $N_2 = 750,000$; $N_k = 500,000$, for $k = 3, 4$; $N_k = 400,000$, for $k = 5, .., 9$; $N_{10} = 250,000$). In each replication of our simulation, we constructed a case-control data set with case-control ratio 1:1 and a total sample size $N = 5000$ from the same cohort. In particular, we sampled $n_k = N_k/1000$ patients ($0.5n_k$ cases and $0.5n_k$ controls) from each study k ($n_1 = 1000$; $n_2 = 750$; $n_k = 500$, for $k = 3, 4$; $n_k = 400$, for $k = 5, 9$; $n_{10} = 250$). Then we analyzed the case-control data set (including K sub-studies) using the proposed IPD and meta-analytical methods under each replication.

We carried out $R = 1,000$ replications under each setting, and summarized the results in terms of relative bias (RB $= (\frac{1}{R}\sum_{r=1}^{R}\hat{\gamma}_{(r)} - \gamma^*)/\gamma^* \times 100\%$), mean of model based variance (MV $= \frac{1}{R}\sum_{r=1}^{R}\hat{\mathbf{v}}(\hat{\gamma}_{(r)})$), empirical variance (EV $= \frac{1}{R-1}\sum_{r=1}^{R}(\hat{\gamma}_{(r)} - \overline{\hat{\gamma}}_{(r)})^2$) and mean squared error (MSE $= \frac{1}{R}\sum_{r=1}^{R}(\hat{\gamma}_{(r)} - \gamma^*)^2$).

Simulation results

The simulation results are shown in Tables 5.1 and 5.2 with G-E independence and dependence holding respectively.

Comparison of methods: As for comparison across methods, we make the following key observations in Tables 5.1 and 5.2: (1) When G and E are independent conditional on $\boldsymbol{S}$, all the methods are unbiased for all the regression parameters; when G and E are dependent given $\boldsymbol{S}$, the bias for the constrained estimator could be very large, especially when G-E association is strong ($c = 0.5$). In this case, the EB estimators controlled the bias by putting less weight on the constrained estimator. EB1 is more robust to this bias compared to EB2 and EB3 especially when G-E association is strong ($c = 0.5$). (2) Regardless of the G-E association, there is a precision gain for both γ_G and γ_{GE} using the constrained estimator as compared to the unconstrained estimator, and the gain is quite dramatic for γ_{GE}. EB1 had greater SE compared to EB2 and EB3 since EB1 is a conservative estimator. The performance of EB2 and EB3 were very similar across all setting in terms of RB, SE and MSE. (3) Given the total sample size $N = 5000$, as comparing the empirical standard errors with the means of the model based standard errors, we found that the proposed variance estimator for each method (except for $\tilde{\boldsymbol{\gamma}}_{\text{IVW-EB}}$) performed well, when G-E association is not very strong ($c = 0, 0.2$). The proposed variance estimator of the EB estimator underestimate the variance when G-E association is strong ($c = 0.5$). (4) when G-E are independent conditional on $\boldsymbol{S}$, the constrained estimator had the smallest MSE; in this case, the estimated τ^2 were very close to 0 in both EB2 and EB3, so EB2 and EB3 showed very similar results to the constrained estimator. When G and E are dependent conditional on $\boldsymbol{S}$, EB 1-3 had smaller MSE as compared to constrained/unconstrained estimators. In particular, when G-E association is modest ($c = 0.2$), EB2 and EB3 had smaller MSE compared to EB1; when G-E association is strong ($c = 0.5$), EB2 and EB3 had greater MSE compared to EB1.

In summary, as for comparison of the methods, when G and E are independent conditional on $\boldsymbol{S}$, all the methods are unbiased, and the constrained estimator has smallest SE.

The EB estimators (especially EB2 and EB3) retain this precision gain by putting more weight on the constrained estimator. When G and E are dependent conditional on $\boldsymbol{S}$, the bias for the constrained estimator could be very large. However, the EB estimators can control for the bias by putting less weight on the constrained estimator while retaining certain amount of precision gain, and show advantage in terms of MSE. We also compared the proposed methods in terms of power and type 1 error. The proposed EB estimators provide increased power compared to a standard case-control analysis, when G and E are independent; and have superior control of type 1 error as compared to a constrained estimator, when G and E are dependent.

Comparison of IPD analysis and meta-analysis: Regardless of the G-E association, we found that the two analysis showed very similar results in terms of RB, SE and MSE, under each of the method. This is expected from our theoretical results. As for $\tilde{\gamma}_{\text{IVW-EB}}$, we observed that it had greater SE and MSE compared to EB1-3, especially when G-E are dependent, implying that pooling the original EB estimator across studies is less efficient than modifying the weights to borrow strength across studies.

5.4 Application: Data analysis for the type 2 diabetes study

The proposed methods were applied to the data from a set of studies investigating T2D, which has been used as the data example in chapter 4 as well. A number of SNPs in the *FTO* gene region (16q12.2) have previously been identified to be associated with T2D and BMI in the DIAGRAM consortium (Zeggini et al. (2008), Voight et al. (2010)). Variants at *FTO* are known to influence T2D predisposition through an effect on BMI (Freathy et al. (2008), Voight et al. (2010)). Age, BMI and gender are all known risk factors for T2D. In this chapter, we investigate whether SNPs in *FTO* gene modifies the effect of environmental factors (e.g. age and BMI) on T2D.

Among the 8 cohorts, the T2D case-control status was identified by the glucose tolerance category, case: fasting glucose$\geq 7.0\ mmol/l$ or two-hour glucose$\geq 11.1\ mmol/l$; control: fasting glucose$< 6.1\ mmol/l$ and two-hour glucose$< 7.8\ mmol/l$. We genotyped all the T2D cases and portions of controls under budget allowance. FUSION-FSIGT and DPS cohort are not valid for a case-control study. The descriptive summary statistics for the genotyped case-control data sets from the remaining 6 cohorts (D2D2007, DIAGEN, FUSION S2, HUNT, METSIM, TROMSO) are shown in Table 4.3, by case-control status. We have a total of $N = 7597$ individuals (3120 cases and 4477 controls), with sample sizes n_k relatively constant but the case-control ratio varying across studies (Table 5.5). The case group has significantly older age (62.7 v.s. 59.2, P-value$<$0.001), higher BMI (29.8 v.s. 26.3, P-value$<$0.001), lower percentage of females (39$\%$ v.s. 47$\%$, P-value$<$0.001) than those in the control group. We examined 9 SNPs in the FTO gene including rs1121980, rs11642841, rs12149832, rs1421085, rs17817449, rs8050136, rs9930506, rs9941349 and rs6499640, which are in high linkage disequilibrium. For each SNP, the MAF ranges 0.4-0.5 and very constant across studies. This is not surprising since the study populations are all European (most are Finnish).

For demonstration purpose, we present our analysis for three representative scenarios: (1) weak G-E association between age and rs11642841 (supporting G-E independence); (2) modest G-E association between BMI and rs1121980; and (3) strong G-E association between BMI and rs6499640 (supporting G-E dependence), which are reflected in the (conditional) G-E association in the control group across the 6 case-control studies. The corresponding G-E correlations were shown in Table 5.5. We explored GEIs on T2D (Y) under the three scenarios. We consider $\boldsymbol{S} = (S_1, S_2)$ as the stratification variable, in which S_1 is age (BMI) when E is BMI (age), and S_2 is gender.

Comparison of methods: The results comparing across methods, under both IPD analysis

and meta-analysis, were shown in Table 5.6. Regardless of the G-E association, we observed a precision gain for γ_{GE} using the constrained or EB estimators, as compared to the unconstrained or logistic regression estimator. In particular, we make the following two key observations: (1) For age×rs11642841 (weak G-E association), all the methods provide very close estimates for all the regression parameters. The constrained model provided the smallest SE; since the estimated τ^2 was 0 in both EB2 and EB3, they showed the same results as the constrained estimator. (2) For BMI×rs1121980 (modest) and BMI×rs6499640 (strong G-E association), the constrained estimate was quite different from the unconstrained estimate for γ_{GE}. The EB estimates (EB1-3) were intermediate, and relatively close to the unconstrain estimates. Moreover, the EB estimators improved the precision substantially as compared to the unconstrained or logistic regression estimator.

Comparison of IPD analysis and meta-analysis: We found that the two analyses showed very consistent results for each of the method in terms of estimates and SEs. As for the IVW EB estimator, we observed that it had greater SE compared to EB2 and EB3 for the GEI parameter γ_{GE}, especially when G-E association is strong.

Results: Under an additive model, with 1 kg/m^2 increase in BMI, the odds ratio of T2D is 1.17 (95% CI: (1.15, 1.19)) given rs6499640=GG, 1.20 (95% CI: (1.17, 1.24)) given rs6499640=AG or GA; and 1.22 (95% CI: (1.19, 1.26)) given rs6499640=AA. The trend of the odds ratios of T2D among the three groups defined by rs6499640 indicated that the presence of minor allele A in rs6499640 enhanced the association between BMI and T2D.

5.5 Discussion

There has been abundance of literature on using G-E independence for case-control studies of G-E interaction. However, there are no papers thus far to study the role of G-

E independence in a meta-analysis setting where the assumption could vary within each study/cohort. In this chapter, we proposed a meta-analysis approach that uses retrospective likelihood as the basis for influence and leverages the G-E independence assumption in a data-adaptive way. The proposed shrinkage estimator provides optimal choices for weights corresponding to constrained and unconstrained models by using information on G-E association parameters derived from multiple studies/cohorts. Our work showed that this novel estimator has better MSE properties than IVW estimator pooling study specific constrained, unconstrained or EB estimators. Our work also lead to many possible extensions such as a mixture distribution prior on the G-E association parameters.

Table 5.1: Comparison of the proposed methods[a] when G-E independence holds, stratified by IPD analysis and meta-analysis.

Method [a]	E				G				GXE			
IPD analysis	RB[b]	SE1[b]	SE2[b]	MSE[b]	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic	-0.001	0.020	0.020	0.412	0.006	0.067	0.065	4.151	0.016	0.025	0.024	0.603
Unconstrained	-0.007	0.018	0.019	0.353	0.014	0.064	0.062	3.816	-0.023	0.023	0.022	0.461
Constrained	-0.004	0.016	0.017	0.276	0.023	0.056	0.054	2.975	-0.036	0.011	0.011	0.127
EB1	-0.007	0.016	0.018	0.310	0.016	0.055	0.058	3.323	-0.026	0.016	0.016	0.254
EB2	-0.004	0.016	0.017	0.276	0.023	0.056	0.054	2.976	-0.036	0.011	0.011	0.128
EB3	-0.004	0.016	0.017	0.276	0.022	0.056	0.054	2.973	-0.036	0.011	0.011	0.127
Meta-analysis	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic (IVW)	0.009	0.020	0.020	0.407	-0.018	0.066	0.064	4.114	-0.011	0.025	0.023	0.534
Unconstrained (IVW)	-0.025	0.018	0.018	0.345	0.035	0.064	0.061	3.766	-0.012	0.024	0.022	0.498
Constrained (IVW)	-0.024	0.017	0.016	0.273	0.014	0.056	0.054	2.931	-0.034	0.012	0.011	0.123
EB1	-0.018	0.015	0.017	0.300	0.033	0.053	0.057	3.284	-0.025	0.015	0.015	0.276
EB2	-0.024	0.017	0.016	0.273	0.015	0.056	0.054	2.927	-0.025	0.012	0.011	0.124
EB3	-0.024	0.017	0.016	0.273	0.015	0.056	0.054	2.931	-0.026	0.012	0.011	0.125
EB (IVW)	-0.030	0.017	0.017	0.294	0.017	0.057	0.056	3.133	-0.013	0.017	0.015	0.329
EB (HIVW)	-0.026	0.017	0.017	0.289	0.016	0.057	0.056	3.081	-0.019	0.016	0.015	0.302

[a] Constrained/unconstrained: the retrospective likelihood method with/without G-E independence assumption. EB1-3: EB estimators of $\hat{\boldsymbol{A}}$ using the following strategies: (1) a conservative estimator $\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^{\top}$; (2) $\hat{\boldsymbol{A}} = diag(\hat{\tau}^2, ..., \hat{\tau}^2)_{(K\times K)}$, where $\hat{\tau}^2$ is the MLE that maximizes the marginal likelihood of $\hat{\boldsymbol{\theta}}$; (3) $\hat{\boldsymbol{A}} = diag(\bar{\tau}^2, ..., \bar{\tau}^2)_{(K\times K)}$, where $\bar{\tau}^2$ is the estimated posterior mean of $\tau^2|\hat{\boldsymbol{\theta}}$. EB (IVW): Inverse variance weighted EB estimator combining EB estimators across study.

[b] RB: relative bias; SE1: mean of model based standard error; SE2: empirical standard error; MSE: mean squared error. (MSEs have been multiplied by 1000.)

Table 5.2: Comparison of the proposed methods[a] when G-E independence is violated, stratified by IPD analysis and meta-analysis.

$\theta_k^* \overset{iid}{\sim}$ Uniform$(-0.2, 0.2)$	E				G				GXE			
IPD analysis	RB[b]	SE1[b]	SE2[b]	MSE[b]	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic	-0.001	0.018	0.019	0.395	0.001	0.066	0.063	3.982	-0.013	0.024	0.025	0.619
Unconstrained	0.020	0.018	0.019	0.428	0.029	0.064	0.062	4.022	-0.020	0.023	0.023	0.595
Constrained	0.062	0.016	0.017	0.419	0.070	0.056	0.054	3.841	-0.177	0.012	0.011	0.630
EB1	0.030	0.016	0.018	0.402	0.037	0.055	0.058	3.603	-0.063	0.015	0.018	0.439
EB2	0.048	0.016	0.017	0.407	0.038	0.055	0.054	3.654	-0.072	0.012	0.012	0.384
EB3	0.053	0.016	0.017	0.416	0.039	0.055	0.054	3.639	-0.071	0.012	0.012	0.381
Meta-analysis	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic(IVW)	0.027	0.019	0.019	0.370	-0.017	0.066	0.062	4.047	-0.026	0.024	0.025	0.629
Unconstrained(IVW)	0.025	0.018	0.018	0.383	0.019	0.064	0.061	4.118	-0.025	0.023	0.023	0.595
Constrained(IVW)	0.071	0.016	0.016	0.387	0.056	0.056	0.054	3.847	-0.246	0.012	0.011	0.727
EB1	0.039	0.015	0.017	0.384	0.036	0.052	0.057	3.635	-0.066	0.015	0.017	0.424
EB2	0.043	0.016	0.016	0.384	0.039	0.055	0.054	3.612	-0.085	0.012	0.011	0.391
EB3	0.042	0.016	0.016	0.383	0.038	0.055	0.054	3.608	-0.084	0.012	0.011	0.392
EB (IVW)	0.071	0.017	0.017	0.364	0.038	0.059	0.057	3.593	-0.066	0.018	0.023	0.573
EB (HIVW)	0.062	0.017	0.017	0.374	0.038	0.057	0.057	3.533	-0.062	0.015	0.019	0.413
$\theta_k^* \overset{iid}{\sim}$ Uniform$(-0.5, 0.5)$	E				G				GXE			
IPD analysis	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic	0.028	0.018	0.017	0.293	-0.040	0.064	0.058	3.462	0.036	0.020	0.019	0.365
Unconstrained	0.041	0.018	0.017	0.300	-0.048	0.063	0.057	3.346	0.011	0.019	0.018	0.353
Constrained	-0.221	0.017	0.015	0.719	-0.245	0.058	0.055	5.358	0.414	0.011	0.010	1.808
EB1	-0.002	0.016	0.017	0.303	-0.073	0.054	0.057	3.446	0.073	0.012	0.019	0.345
EB2	-0.112	0.016	0.017	0.408	-0.153	0.058	0.056	4.014	0.240	0.012	0.016	0.616
EB3	-0.105	0.016	0.017	0.396	-0.147	0.058	0.056	3.960	0.229	0.012	0.016	0.675
Meta-analysis	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic(IVW)	0.015	0.018	0.017	0.277	-0.042	0.065	0.057	3.278	-0.023	0.020	0.019	0.351
Unconstrained(IVW)	0.023	0.018	0.017	0.278	-0.035	0.063	0.056	3.219	-0.033	0.020	0.018	0.344
Constrained(IVW)	0.008	0.017	0.015	0.217	-0.202	0.058	0.053	4.436	0.563	0.014	0.010	3.275
EB1	0.023	0.016	0.016	0.272	-0.042	0.056	0.056	3.206	-0.002	0.014	0.018	0.323
EB2	0.022	0.017	0.016	0.250	-0.082	0.058	0.055	3.245	0.160	0.013	0.018	0.567
EB3	0.023	0.017	0.016	0.252	-0.078	0.058	0.055	3.233	0.143	0.013	0.018	0.522
EB (IVW)	0.091	0.018	0.017	0.376	-0.013	0.063	0.060	3.597	-0.153	0.019	0.022	0.709
EB (HIVW)	0.067	0.017	0.017	0.296	-0.012	0.059	0.060	3.467	-0.147	0.017	0.019	0.653

[a] Constrained/unconstrained: the retrospective likelihood method with/without G-E independence assumption. EB1-3: EB estimators of $\hat{\boldsymbol{A}}$ using the following strategies: (1) a conservative estimator $\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top$; (2) $\hat{\boldsymbol{A}} = diag(\hat{\tau}^2, ..., \hat{\tau}^2)_{(K\times K)}$, where $\hat{\tau}^2$ is the MLE that maximizes the marginal likelihood of $\hat{\boldsymbol{\theta}}$; (3) $\hat{\boldsymbol{A}} = diag(\bar{\tau}^2, ..., \bar{\tau}^2)_{(K\times K)}$, where $\bar{\tau}^2$ is the estimated posterior mean of $\tau^2|\hat{\boldsymbol{\theta}}$. EB (IVW): Inverse variance weighted EB estimator combining EB estimators across study.

[b] RB: relative bias; SE1: mean of model based standard error; SE2: empirical standard error; MSE: mean squared error. (MSEs have been multiplied by 1000.)

Table 5.3: Comparison of the proposed methods[a] when G-E independence is violated (G-E associations follow a mixture distribution), stratified by IPD analysis and meta-analysis.

$\theta_k^{*iid}\sim 0.4\delta(0)+0.6\text{U}(-0.3,0.3)$	E				G				GXE			
IPD analysis	RB[b]	SE1[b]	SE2[b]	MSE[b]	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic	-0.013	0.018	0.018	0.336	0.016	0.064	0.065	4.317	0.019	0.022	0.022	0.489
Unconstrained	-0.004	0.018	0.018	0.322	0.015	0.063	0.063	3.944	0.004	0.021	0.021	0.431
Constrained	-0.058	0.017	0.017	0.333	0.027	0.056	0.055	3.185	0.055	0.011	0.011	0.136
EB1	-0.036	0.015	0.017	0.317	0.026	0.054	0.058	3.515	0.017	0.015	0.016	0.250
EB2	-0.053	0.016	0.017	0.325	0.028	0.056	0.056	3.232	0.030	0.011	0.012	0.138
EB3	-0.052	0.016	0.017	0.324	0.028	0.056	0.056	3.239	0.029	0.011	0.012	0.140
Meta-analysis	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic(IVW)	-0.027	0.018	0.018	0.326	0.010	0.065	0.064	4.249	-0.052	0.022	0.021	0.478
Unconstrained(IVW)	-0.025	0.018	0.018	0.315	0.020	0.063	0.062	3.904	-0.047	0.021	0.020	0.434
Constrained(IVW)	0.008	0.017	0.017	0.279	0.044	0.056	0.054	3.115	0.101	0.012	0.011	0.214
EB1	-0.005	0.015	0.017	0.287	0.034	0.050	0.058	3.551	-0.013	0.014	0.017	0.302
EB2	0.007	0.016	0.017	0.278	0.034	0.056	0.055	3.218	0.060	0.012	0.012	0.182
EB3	0.007	0.016	0.017	0.278	0.035	0.056	0.055	3.231	0.056	0.012	0.012	0.183
EB (IVW)	0.031	0.017	0.017	0.309	0.078	0.061	0.061	4.311	-0.138	0.018	0.021	0.617
EB (HIVW)	0.026	0.016	0.017	0.289	0.064	0.059	0.060	3.957	-0.103	0.016	0.018	0.412
$\theta_k^{*iid}\sim 0.4\delta(0)+0.6\text{U}(-0.5,0.5)$	E				G				GXE			
IPD analysis	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic	0.008	0.018	0.017	0.297	-0.020	0.064	0.062	3.749	0.031	0.021	0.021	0.454
Unconstrained	0.040	0.018	0.017	0.310	-0.022	0.063	0.060	3.633	-0.027	0.020	0.020	0.407
Constrained	-0.107	0.017	0.016	0.380	-0.096	0.057	0.054	3.262	0.284	0.012	0.011	0.593
EB1	-0.007	0.016	0.017	0.302	-0.027	0.052	0.058	3.351	0.029	0.014	0.018	0.334
EB2	-0.069	0.016	0.017	0.324	-0.064	0.057	0.055	3.175	0.122	0.012	0.013	0.312
EB3	-0.066	0.016	0.017	0.321	-0.061	0.057	0.055	3.179	0.116	0.012	0.013	0.309
Meta-analysis	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE	RB	SE1	SE2	MSE
Standard logistic(IVW)	-0.006	0.018	0.017	0.279	-0.011	0.065	0.061	3.811	-0.031	0.021	0.021	0.448
Unconstrained(IVW)	0.019	0.018	0.017	0.280	-0.011	0.063	0.060	3.630	-0.031	0.020	0.020	0.460
Constrained(IVW)	0.080	0.017	0.016	0.315	-0.022	0.058	0.053	2.818	0.240	0.013	0.010	0.673
EB1	0.033	0.016	0.016	0.274	-0.002	0.054	0.058	3.382	-0.037	0.014	0.020	0.400
EB2	0.064	0.017	0.016	0.287	0.003	0.057	0.055	2.979	0.087	0.012	0.016	0.322
EB3	0.062	0.017	0.016	0.285	0.003	0.057	0.055	3.003	0.076	0.012	0.016	0.320
EB (IVW)	0.060	0.017	0.016	0.290	-0.002	0.061	0.059	3.438	-0.130	0.018	0.021	0.594
EB (HIVW)	0.060	0.017	0.016	0.288	-0.001	0.060	0.059	3.295	-0.117	0.018	0.020	0.578

[a] Constrained/unconstrained: the retrospective likelihood method with/without G-E independence assumption. EB1-3: EB estimators of $\hat{\boldsymbol{A}}$ using the following strategies: (1) a conservative estimator $\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top$; (2) $\hat{\boldsymbol{A}}=diag(\hat{\tau}^2,...,\hat{\tau}^2)_{(K\times K)}$, where $\hat{\tau}^2$ is the MLE that maximizes the marginal likelihood of $\hat{\boldsymbol{\theta}}$; (3) $\hat{\boldsymbol{A}}=diag(\bar{\tau}^2,...,\bar{\tau}^2)_{(K\times K)}$, where $\bar{\tau}^2$ is the estimated posterior mean of $\tau^2|\hat{\boldsymbol{\theta}}$. EB (IVW): Inverse variance weighted EB estimator combining EB estimators across study.

[b] RB: relative bias; SE1: mean of model based standard error; SE2: empirical standard error; MSE: mean squared error. (MSEs have been multiplied by 1000.)

Table 5.4: Comparison of the proposed IPD/meta-analytical methods in terms of power and Type-I error, where E is standardized.

	Power ($\gamma^*_{GE} = 0.2$)		Type-I error ($\gamma^*_{GE} = 0$)	
IPD analysis	$\theta^*_k = 0$	$\theta^*_k \overset{iid}{\sim} \text{U}(-0.5, 0.5)$	$\theta^*_k = 0$	$\theta^*_k \overset{iid}{\sim} \text{U}(-0.5, 0.5)$
Standard logistic	0.69	0.70	0.05	0.05
Unconstrained	0.71	0.72	0.05	0.06
Constrained	0.92	0.95	0.06	0.38
EB1	0.81	0.78	0.05	0.07
EB2	0.84	0.81	0.06	0.11
EB3	0.82	0.80	0.06	0.12
Meta-analysis				
Standard logistic	0.68	0.69	0.05	0.05
Unconstrained	0.70	0.72	0.05	0.06
Constrained	0.92	0.94	0.06	0.35
EB1	0.79	0.78	0.05	0.06
EB2	0.81	0.82	0.06	0.10
EB3	0.81	0.81	0.05	0.12
EB (IVW)	0.77	0.81	0.05	0.14
EB (HIVW)	0.80	0.82	0.05	0.11

Table 5.5: Summary statistics for the 6 case-control studies.

		Age (year)		BMI[a](kg/m^2)		Female	rs11642841			rs6499640			rs1121980		
Study	N	mean	SD[a]	mean	SD	ratio	MAF[a]	(age	P)[b]	MAF	(BMI	P)	MAF	(BMI	P)
D2D2007	1330	58.3	8.3	26.9	4.9	0.60	0.41	-0.02	0.53	0.43	-0.06	0.02^{*}	0.40	0.04	0.15
DIAGEN	1043	61.1	14.1	28.1	5.4	0.57	0.42	<0.01	0.96	0.40	0.01	0.69	0.47	0.03	0.34
FUSION S2	1418	60.5	8.1	28.3	5.1	0.45	0.40	-0.01	0.74	0.43	-0.04	0.16	0.39	0.06	0.03^{*}
HUNT	1232	67.3	13.1	27.9	4.4	0.49	0.45	-0.04	0.21	0.37	-0.03	0.33	0.47	0.06	0.04^{*}
METSIM	1235	56.7	6.8	28.0	4.8	0	0.42	-0.02	0.56	0.41	-0.02	0.41	0.44	0.03	0.28
TROMSO	1337	60.1	12.4	27.5	4.7	0.51	0.45	0.02	0.55	0.38	-0.01	0.78	0.49	0.05	0.07
Total	7595	**60.6**	11.2	**27.8**	4.9	**0.43**	**0.43**	<0.01	0.91	**0.40**	-0.03	0.03^{*}	**0.44**	0.04	0.01^{***}
D2D2007	287	62.3	7.8	30.2	5.6	0.44	0.42	<0.01	0.98	0.39	0.02	0.78	0.40	0.11	0.06
DIAGEN	421	66.2	11.6	30.1	6.2	0.50	0.44	<0.01	0.96	0.40	0.05	0.36	0.48	0.05	0.35
FUSION S2	624	60.9	8.3	30.5	5.4	0.46	0.42	-0.07	0.10	0.42	0.05	0.26	0.41	0.04	0.30
HUNT	511	69.1	11.3	29.1	4.7	0.49	0.48	-0.09	0.04^{*}	0.38	<0.01	0.92	0.50	0.05	0.23
METSIM	632	59.7	6.9	30.0	5.3	0	0.43	-0.03	0.52	0.40	-0.02	0.63	0.44	0.03	0.44
TROMSO	644	60.2	12.4	29.2	4.9	0.51	0.45	0.04	0.33	0.37	-0.01	0.87	0.50	0.09	0.03^{*}
Total cases	3119	**62.7**	10.6	**29.8**	5.4	**0.39**	**0.44**	-0.01	0.66	**0.39**	0.02	0.34	**0.46**	0.05	0.01^{***}
D2D2007	1043	57.3	8.1	26.0	4.3	0.64	0.41	-0.03	0.37	0.44	-0.07	0.03^{*}	0.40	0.02	0.51
DIAGEN	622	57.7	14.6	26.7	4.4	0.61	0.41	-0.02	0.68	0.39	-0.03	0.49	0.46	<0.01	0.95
FUSION S2	794	60.2	8.0	26.6	4.0	0.44	0.38	0.03	0.33	0.43	-0.11	0.01^{**}	0.37	0.03	0.40
HUNT	721	66.0	14.1	27.0	4.0	0.48	0.44	-0.01	0.71	0.37	-0.05	0.14	0.45	0.04	0.32
METSIM	603	53.5	5.0	25.8	3.0	0	0.41	-0.04	0.31	0.43	0.02	0.69	0.45	0.06	0.12
TROMSO	693	60.0	12.5	25.9	3.8	0.50	0.44	<0.01	0.93	0.38	<0.01	0.95	0.49	<0.01	0.99
Total controls	4476	**59.2**	11.3	**26.3**	4.0	**0.47**	**0.41**	<0.01	0.92	**0.41**	-0.05	0.01^{***}	**0.43**	0.02	0.16

[a] SD: standard deviation; BMI: body mass index; MAF: minor allele frequency.

[b] Correlation(P): Spearman correlation between SNP and environmental factor (age and BMI) with corresponding P-value.

* $* < 0.05$; $** < 0.01$; $*** < 0.001$.

Table 5.6: Results comparing the proposed IPD and meta-analytical methods under different scenarios of G-E dependence/independence for the T2D study, where we used type 2 diabetes for Y, SNPs on FTO gene for G, either age or BMI for E (the other for S_1) and gender for S_2.

Methods [a]	age***				rs11642841*				age×rs11642841			
IPD analysis	Est[b]	SE[b]	95% CI[b]		Est	SE	95% CI		Est	SE	95% CI	
Standard logistic	0.398	0.039	0.321	0.475	0.733	0.369	0.009	1.456	-0.025	0.034	-0.091	0.042
Unconstrained	0.403	0.037	0.329	0.476	0.731	0.370	0.005	1.457	-0.020	0.031	-0.081	0.041
Constrained	0.410	0.034	0.343	0.477	0.724	0.369	0.001	1.448	-0.028	0.025	-0.077	0.021
EB1	0.410	0.035	0.340	0.479	0.725	0.347	0.044	1.406	-0.027	0.029	-0.084	0.029
EB2	0.410	0.034	0.343	0.477	0.724	0.369	0.001	1.448	-0.028	0.025	-0.077	0.021
EB3	0.410	0.034	0.343	0.477	0.724	0.369	0.001	1.448	-0.028	0.025	-0.077	0.021
Meta-analysis	Est	SE	95% CI		Est	SE	95% CI		Est	SE	95% CI	
Standard logistic	0.358	0.039	0.280	0.435	0.789	0.379	0.047	1.531	-0.026	0.034	-0.093	0.040
Unconstrained	0.352	0.038	0.278	0.426	0.753	0.378	0.013	1.493	-0.019	0.031	-0.080	0.042
Constrained	0.357	0.034	0.290	0.424	0.724	0.371	-0.002	1.451	-0.027	0.025	-0.075	0.022
EB1	0.357	0.036	0.286	0.427	0.726	0.348	0.045	1.407	-0.026	0.030	-0.084	0.032
EB2	0.357	0.034	0.290	0.424	0.724	0.371	-0.002	1.451	-0.027	0.025	-0.075	0.022
EB3	0.357	0.034	0.290	0.424	0.724	0.371	-0.002	1.451	-0.027	0.025	-0.075	0.022
EB (IVW)	0.350	0.035	0.280	0.419	0.787	0.376	0.051	1.524	-0.016	0.028	-0.070	0.038
EB (HIVW)	0.353	0.035	0.282	0.422	0.767	0.374	0.042	1.484	-0.024	0.029	-0.081	0.033
	BMI***				rs6499640				BMI×rs6499640*			
IPD analysis	Est	SE	95% CI		Est	SE	95% CI		Est	SE	95% CI	
Standard logistic	1.568	0.093	1.386	1.750	-0.225	0.378	-0.967	0.516	0.229	0.090	0.053	0.405
Unconstrained	1.590	0.085	1.424	1.756	-0.216	0.374	-0.949	0.517	0.202	0.074	0.057	0.347
Constrained	1.735	0.073	1.591	1.878	-0.484	0.365	-1.200	0.232	0.018	0.048	-0.076	0.113
EB1	1.608	0.068	1.474	1.742	-0.251	0.344	-0.924	0.423	0.179	0.052	0.076	0.281
EB2	1.609	0.078	1.456	1.761	-0.252	0.385	-1.006	0.502	0.178	0.061	0.059	0.297
EB3	1.603	0.079	1.448	1.758	-0.241	0.386	-0.998	0.516	0.185	0.064	0.061	0.310
Meta-analysis	Est	SE	95% CI		Est	SE	95% CI		Est	SE	95% CI	
Standard logistic	1.452	0.093	1.268	1.635	-0.147	0.387	-0.906	0.612	0.222	0.091	0.044	0.400
Unconstrained	1.479	0.085	1.311	1.646	-0.160	0.381	-0.906	0.586	0.202	0.074	0.056	0.348
Constrained	1.630	0.074	1.485	1.775	-0.447	0.372	-1.175	0.282	0.015	0.048	-0.080	0.110
EB1	1.498	0.070	1.361	1.634	-0.196	0.350	-0.882	0.490	0.179	0.053	0.076	0.282
EB2	1.499	0.078	1.345	1.652	-0.198	0.392	-0.966	0.571	0.177	0.061	0.057	0.297
EB3	1.493	0.080	1.336	1.649	-0.186	0.394	-0.958	0.586	0.185	0.064	0.060	0.310
EB (IVW)	1.572	0.077	1.421	1.724	-0.440	0.375	-1.174	0.294	0.107	0.067	-0.026	0.239
EB (HIVW)	1.543	0.077	1.416	1.718	-0.375	0.381	-1.012	0.286	0.147	0.060	0.029	0.264
	BMI***				rs1121980				BMI×rs1121980*			
IPD analysis	Est	SE	95% CI		Est	SE	95% CI		Est	SE	95% CI	
Standard logistic	1.703	0.100	1.507	1.898	0.476	0.370	-0.250	1.201	0.049	0.087	-0.121	0.218
Unconstrained	1.691	0.090	1.514	1.868	0.466	0.368	-0.256	1.188	0.061	0.073	-0.081	0.204
Constrained	1.621	0.076	1.472	1.769	0.593	0.360	-0.113	1.300	0.142	0.047	0.049	0.235
EB1	1.661	0.071	1.522	1.801	0.520	0.313	-0.094	1.134	0.095	0.050	-0.002	0.193
EB2	1.658	0.074	1.514	1.802	0.525	0.369	-0.197	1.248	0.099	0.043	0.016	0.182
EB3	1.665	0.075	1.517	1.813	0.513	0.372	-0.216	1.241	0.091	0.047	0.001	0.182
Meta-analysis	Est	SE	95% CI		Est	SE	95% CI		Est	SE	95% CI	
Standard logistic	1.578	0.101	1.379	1.777	0.578	0.378	-0.164	1.320	0.060	0.088	-0.112	0.232
Unconstrained	1.579	0.091	1.400	1.757	0.563	0.375	-0.172	1.298	0.064	0.073	-0.080	0.207
Constrained	1.514	0.077	1.364	1.664	0.662	0.367	-0.057	1.381	0.138	0.048	0.045	0.232
EB1	1.548	0.071	1.408	1.688	0.609	0.315	-0.008	1.227	0.099	0.049	0.003	0.194
EB2	1.543	0.074	1.398	1.687	0.618	0.372	-0.112	1.347	0.105	0.040	0.026	0.184
EB3	1.549	0.075	1.402	1.696	0.608	0.374	-0.126	1.342	0.098	0.043	0.013	0.182
EB (IVW)	1.511	0.086	1.343	1.679	0.662	0.370	-0.064	1.388	0.137	0.065	0.009	0.264
EB (HIVW)	1.526	0.082	1.356	1.677	0.632	0.371	-0.043	1.418	0.142	0.052	0.029	0.262

[a] Constrained/unconstrained: the retrospective likelihood method with/without G-E independence assumption. EB1-3: EB estimators of $\hat{\boldsymbol{A}}$ using the following strategies: (1) a conservative estimator $\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top$; (2) $\hat{\boldsymbol{A}} = diag(\hat{\tau}^2, ..., \hat{\tau}^2)_{(K\times K)}$, where $\hat{\tau}^2$ is the MLE that maximizes the marginal likelihood of $\hat{\boldsymbol{\theta}}$; (3) $\hat{\boldsymbol{A}} = diag(\bar{\tau}^2, ..., \bar{\tau}^2)_{(K\times K)}$, where $\bar{\tau}^2$ is the estimated posterior mean of $\tau^2|\hat{\boldsymbol{\theta}}$. EB (IVW): Inverse variance weighted EB estimator combining EB estimators across study.

[b] Est: estimate; SE: standard error; CI: confidence interval. (All numbers have been multiplied by 10 to compare the results in 3 significant digits.)

[*] $* < 0.05$; $** < 0.01$; $*** < 0.001$.

CHAPTER VI

Conclusion

My dissertation work is around the theme of Bayesian modeling for environmental association and gene-environment interaction under complex epidemiologic study designs. The first two projects considered the problems of characterizing effect of environmental exposure on health outcomes under complex sampling designs, in particular, a matched case-control study with multiple disease sub-types and nonlinear odds ratio functions for point source modeling in project 2 and Bayesian analysis of time-series data under case-crossover designs under project 3. In the last two projects, we addressed several important issues in meta-analysis of G-E interactions. We studied the role of G-E independence and environmental heterogeneity across studies on the characteristics of several meta-analysis approaches. We focused on quantitative traits in project 3 and dichotomous traits under case-control sampling in the last project. The following paragraphs list the corresponding conclusion remarks for each chapter respectively.

In the first project, we considered sub-types within cases under a matched case-control design for point sources modeling with nonlinear odds ratio functions. Frequentist and Bayesian inference methods were evaluated. With multiple parameters and non-linear model, Bayesian methods using MCMC techniques appear to have advantages in terms of estimation stability and precision over the frequentist alternatives. Moreover, the posterior

estimate of the odds ratio function $\hat{\boldsymbol{f}}_{\boldsymbol{k}}$ and the corresponding HPD interval can be obtained based on exact posterior draws that avoids large sample approximation such as the Delta theorem for a frequentist MLE. For a MLE, the asymptotic properties might not hold for the complex likelihood derived from the non-linear models, especially when the sample size is not large. The proposed methods were applied to a population-based matched case-control study investigating associations between acute asthma outcomes and proximity of residence to major roads by analyzing Medicaid claims data for the pediatric asthma population in Detroit, MI, from 2004-2006. We addressed the research problem regarding associations between acute asthma risk and proximity of residence to major roads, and showed that the proposed PCM offered better fit compared to the model with standard case-control status. The results indicated that the odds ratio at the point source is 1.2-1.3 as compared to the background odds of disease in the case-control population, and the freeway effect on asthma lasts up to around 500 meters where the first 250 meters have stronger influence.

The second project considered Bayesian analysis of time-series data under case-crossover designs. The contribution of this project is two-fold. First, we established Bayesian equivalence results that require characterization of the time function S_t in a log-linear model for a given choice of window $W(t)$ under a particular case-crossover design, and require characterization of the set of priors under which the posterior distributions of the risk ratio parameters based on a case-crossover design and log-linear model are identical. Second, we considered a more general full likelihood-based approach which made less restrictive assumptions on the baseline risk models and exposure series. We proposed a semi-parametric Bayesian approach using a Dirichlet process prior to handle the random nuisance parameters that appear in a full likelihood formulation under a case-crossover design. This work leads to many potential extensions where a Bayesian analysis may have

attractive features, such as distributed lag linear/non-linear models and hierarchical models for meta-analyses. The proposed methods were illustrated through the DAMAT study, but instead of the distance to the major roads, we focused on the effect of ambient air pollutant concentrations on the acute asthma risk. Evidence of significant increases in acute asthma risk was found with 10 $\mu g\, m^{-3}$ increase in $PM_{2.5}$ concentrations leading to a risk ratio ranging from 1.02 to 1.06 across different methods.

The third project considered meta-analysis of G-E interaction for quantitative traits. In this project, we studied the effect of environmental covariate heterogeneity (within and between cohorts) on two approaches for fixed-effects meta-analysis: the standard inverse-variance weighted meta-analysis and a meta-regression approach. Akin to the results obtained in Simmonds and Higgins (2007), we obtain analytical efficiency expressions for both methods under the assumption of gene-environment independence. The relative efficiency of the two methods depend on the ratio of within versus between cohort variance of the environmental covariate. We propose to use an adaptively weighted estimator (AWE), as a combination of meta-analysis and meta-regression estimators, that can be used as a default choice, retaining full efficiency of the 'gold standard' joint analysis for the interaction parameter using individual level data under certain natural assumptions. The AWE improves efficiency by combining meta-analysis and meta-regression based on only univariate summary statistics from each study, and bypasses issues with sharing of individual level data or multivariate information matrices across studies without sacrificing efficiency. AWE also has advantage over MIVW when the effect of G or E are uncommon across studies. We compared the performance of the proposed methods under a wide spectrum of scenarios and showed that the AWE can serve as a default estimator. The methods were illustrated through meta-analysis of GEI between SNP in the *FTO* gene and BMI on HDL-C data from a set of T2D studies. Under an additive model, with 1 kg/m^2

increase in BMI, HDL-C level on average decreased by 1.73% (95% CI: (1.57, 1.90)%) given rs1121980=GG, by 1.54% (95% CI: (1.44, 1.64)%) given rs1121980=AG or GA; and by 1.35% (95% CI: (1.17, 1.53)%) given rs1121980=AA. The results indicated that the presence of minor allele A in rs1121980 attenuated the negative association between BMI and HDL-C.

The last project extended the work of project 3 to dichotomous traits under case-control studies. In this project, we proposed a meta-analysis approach that uses retrospective likelihood as the basis for inference and leverages the G-E independence assumption in a data-adaptive way. The proposed shrinkage estimator provides optimal choices for weights corresponding to constrained and unconstrained models by using information on G-E association parameters derived from multiple studies/cohorts. Our work showed that this novel estimator has better MSE properties than IVW estimator pooling study specific constrained, unconstrained or EB estimators. The results were illustrated through the T2D studies as well. We used T2D status as the case-control outcome and studied the GEI between SNPs in the FTO gene and environmental factors such as age and BMI for demonstration purpose. Under an additive model, with 1 kg/m^2 increase in BMI, the odds ratio of T2D is 1.17 (95% CI: (1.15, 1.19)) given rs6499640=GG, 1.20 (95% CI: (1.17, 1.24)) given rs6499640=AG or GA; and 1.22 (95% CI: (1.19, 1.26)) given rs6499640=AA. The trend of the odds ratios of T2D among the three groups defined by rs6499640 indicated that the presence of minor allele A in rs6499640 enhanced the association between BMI and T2D.

In summary, my dissertation work is expected to contribute to important analytical and methodological issues that have relevance and applications in the area of genetic and environmental epidemiology.

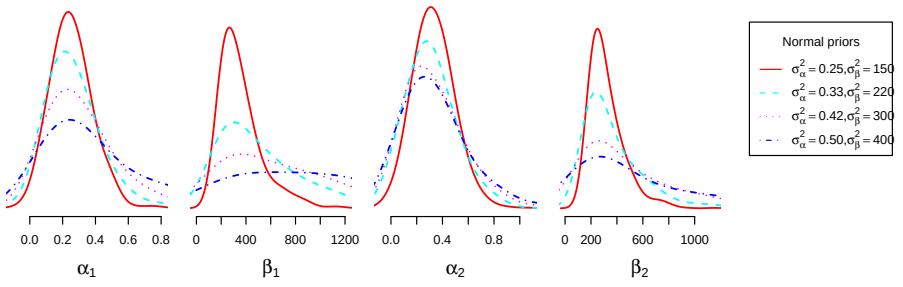
APPENDICES

APPENDIX A

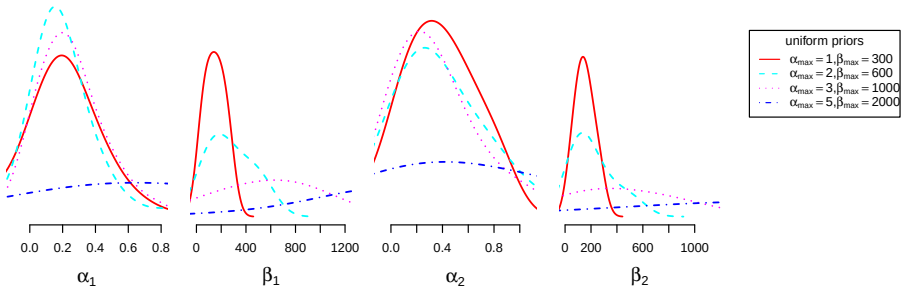
(Appendix Figures and Tables)

A.1 Supplementary Figures and Tables for Chapter 2

Figure A.1: Estimated posterior densities for different settings of prior choices for the **one point source polychotomous category model** for the Detroit Medicaid data, as a sensitivity analysis.

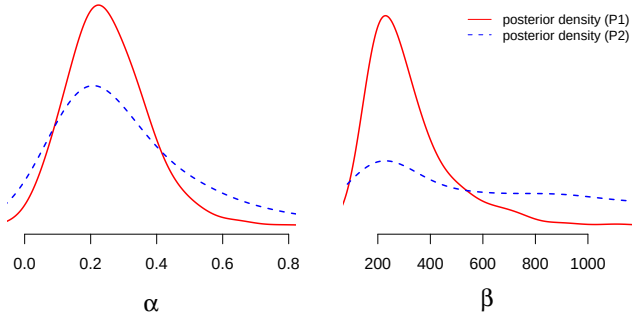


(a) Estimated posterior densities using normal priors on $\log(1+\alpha)$ and $\log(\beta)$

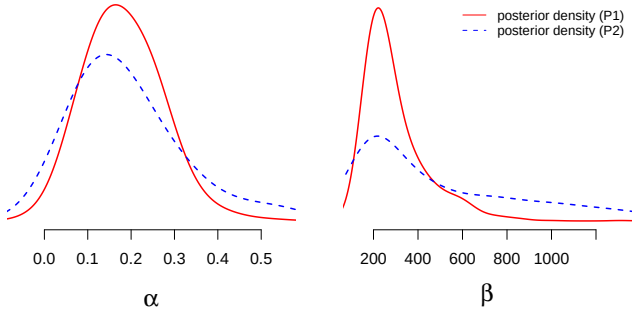


(b) Estimated posterior densities using uniform priors on $(-1, \alpha_{max})$ and $(-1, \beta_{max})$

Figure A.2: Estimated posterior densities for different settings of prior choices for the **one point source binary model and homogeneous adjacent category model** for the Detroit Medicaid data. Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$.

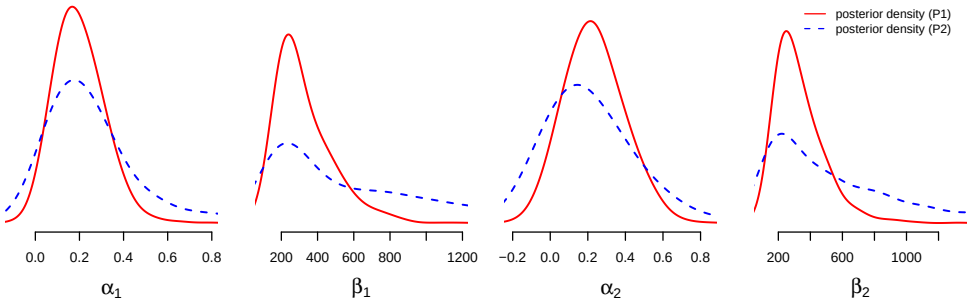


(a) one point source binary model

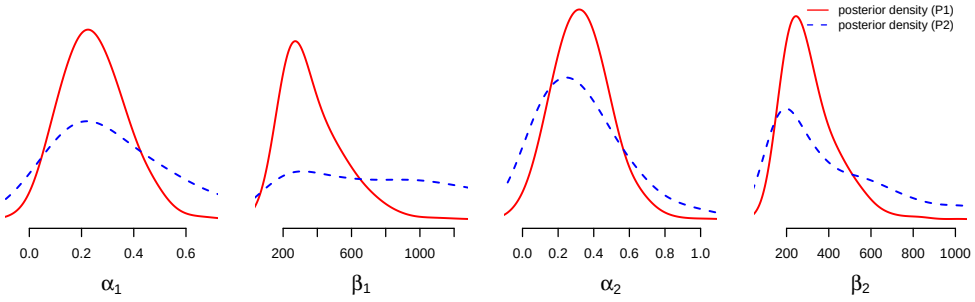


(b) one point source homogeneous adjacent category model

Figure A.3: Estimated posterior densities for different settings of prior choices for the **one point source adjacent category model and polychotomous category model** for the Detroit Medicaid data. Prior 1: $(\mu_{\alpha_1}, \mu_{\beta_1}) = (\mu_{\alpha_2}, \mu_{\beta_2}) = (0.5, 400)$ and $(\sigma^2_{\alpha_1}, \sigma^2_{\beta_1}) = (\sigma^2_{\alpha_2}, \sigma^2_{\beta_2}) = (0.25, 150)$; Prior 2: $(\mu_{\alpha_1}, \mu_{\beta_1}) = (\mu_{\alpha_2}, \mu_{\beta_2}) = (0.5, 400)$ and $(\sigma^2_{\alpha_1}, \sigma^2_{\beta_1}) = (\sigma^2_{\alpha_2}, \sigma^2_{\beta_2}) = (0.5, 400)$.



(a) one point source adjacent category model



(b) one point source polychotomous category model

Table A.1: Summary statistics across the five inference methods for the **one point source adjacent category model (homogeneousness)**, based on $R = 500$ simulations with sample size $\boldsymbol{N = 500}$.

Sample size N=500		$u = log(1+\alpha)$			$v = log(\beta)$		
(α, β)	Method[a]	RB[b](%)	MSE[b]	CP[b](%)	RB(%)	MSE	CP(%)
(0.7, 500)	MLE	19.6	0.036	97	1.7	0.113	92
	Profile	12.9	0.024	98	0.3	0.089	90
	IRLS	13.0	0.024	98	0.2	0.088	90
	Bayesian P1	-10.3	0.006	98	-0.6	0.017	99
	Bayesian P2	5.7	0.008	98	1.2	0.055	98
(0.7, 300)	MLE	9.4	0.018	95	0.9	0.087	90
	Profile	9.0	0.017	98	-0.8	0.076	91
	IRLS	9.4	0.018	98	-0.8	0.076	90
	Bayesian P1	-14.2	0.010	93	3.3	0.049	94
	Bayesian P2	-1.4	0.010	99	2.0	0.055	93
(0.4, 500)	MLE	24.8	0.021	96	-0.2	0.228	80
	Profile	24.7	0.023	98	-0.2	0.218	92
	IRLS	24.8	0.021	98	-0.2	0.199	92
	Bayesian P1	4.4	0.003	100	-0.8	0.016	99
	Bayesian P2	17.6	0.009	95	2.7	0.112	98
(0.4, 300)	MLE	25.6	0.025	96	-0.3	0.196	87
	Profile	25.3	0.025	96	-0.3	0.196	88
	IRLS	25.6	0.025	96	-0.3	0.196	88
	Bayesian P1	4.1	0.003	100	4.5	0.080	97
	Bayesian P2	13.4	0.008	96	4.9	0.204	96
(0.5, 200)	MLE	16.2	0.026	98	-0.3	0.192	81
	Profile	16.3	0.026	99	-0.3	0.192	89
	IRLS	16.2	0.026	99	-0.2	0.192	90
	Bayesian P1	-15.8	0.007	96	5.3	0.321	91
	Bayesian P2	-5.1	0.007	99	6.0	0.539	86

[a] MLE: Maximum likelihood estimate; Profile: Profile likelihood based estimate and confidence interval; IRLS: iteratively re-weighted least squares; Bayesian P1 and P2 refer to two settings of prior choice; Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$.

[b] RB: relative bias; MSE: mean squared error; CP: coverage probability.

Table A.2: Summary statistics across the five inference methods for the **one point source adjacent category model**, based on $R = 500$ simulations with sample size $\boldsymbol{N = 1000}$.

Sample size N=1000		$u_{11} = log(1+\alpha_{11})$			$v_{11} = log(\beta_{11})$			$u_{12} = log(1+\alpha_{12})$			$v_{12} = log(\beta_{12})$		
$(\alpha_{11}, \beta_{11}, \alpha_{12}, \beta_{12})$	Method[a]	RB[b](%)	MSE[b]	CP[b](%)	RB(%)	MSE	CP(%)	RB(%)	MSE	CP(%)	RB(%)	MSE	CP(%)
(0.4, 400,0.7, 800)	MLE	7.9	0.005	96	0.5	0.050	92	17.0	0.055	99	0.9	0.095	95
	Profile	8.0	0.006	95	0.5	0.057	92	17.5	0.057	99	0.9	0.108	94
	IRLS	7.9	0.006	95	0.5	0.053	92	16.9	0.050	99	0.9	0.101	93
	Bayesian P1	9.2	0.003	96	2.7	0.045	93	-20.1	0.014	90	-4.4	0.103	89
	Bayesian P2	3.4	0.005	94	1.5	0.040	95	13.0	0.011	99	2.4	0.089	97
(0.3, 300,0.7, 500)	MLE	6.9	0.005	97	1.4	0.120	89	4.8	0.012	95	0.2	0.047	97
	Profile	6.5	0.006	96	0.8	0.120	92	4.8	0.013	95	0.4	0.053	93
	IRLS	6.3	0.005	96	0.8	0.108	92	4.9	0.013	95	0.4	0.048	93
	Bayesian P1	14.1	0.003	95	4.5	0.086	89	-13.6	0.008	94	-0.9	0.017	100
	Bayesian P2	7.5	0.004	97	2.1	0.062	95	6.7	0.008	97	0.7	0.041	97
(0.5, 500,0.5, 500)	MLE	3.5	0.004	97	-0.6	0.038	94	15.6	0.016	99	1.4	0.123	92
	Profile	3.4	0.005	97	-0.4	0.045	91	15.1	0.017	99	0.7	0.135	90
	IRLS	3.3	0.004	97	-0.4	0.041	92	15.4	0.016	98	0.7	0.124	89
	Bayesian P1	1.9	0.002	99	-0.5	0.013	99	1.6	0.003	100	-0.4	0.018	100
	Bayesian P2	6.9	0.004	97	0.0	0.026	98	15.7	0.010	99	2.6	0.116	98
(0.7, 800,0.5, 400)	MLE	11.1	0.007	94	0.8	0.191	84	4.5	0.010	98	-0.2	0.121	89
	Profile	12.4	0.009	92	0.3	0.224	86	5.1	0.012	96	-0.2	0.135	88
	IRLS	12.0	0.008	92	0.2	0.206	86	4.9	0.011	96	-0.2	0.123	88
	Bayesian P1	1.2	0.002	98	8.3	0.226	67	-10.4	0.004	99	-0.4	0.022	99
	Bayesian P2	9.5	0.005	96	2.0	0.077	96	4.4	0.005	98	1.4	0.081	96

[a] MLE: Maximum likelihood estimate; Profile: Profile likelihood based estimate and confidence interval; IRLS: iteratively re-weighted least squares; Bayesian P1 and P2 refer to two settings of prior choice; Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$.

[b] RB: relative bias; MSE: mean squared error; CP: coverage probability.

Table A.3: Summary statistics across the five inference methods for the **one point source polychotomous category model**, based on $R = 500$ simulations with sample size $\boldsymbol{N = 1000}$.

Sample size N=1000		$u_{11} = log(1+\alpha_{11})$			$v_{11} = log(\beta_{11})$			$u_{12} = log(1+\alpha_{12})$			$v_{12} = log(\beta_{12})$		
$(\alpha_{11}, \beta_{11}, \alpha_{12}, \beta_{12})$	Method[a]	RB[b](%)	MSE[b]	CP[b](%)	RB(%)	MSE	CP(%)	RB(%)	MSE	CP(%)	RB(%)	MSE	CP(%)
(0.4, 400, 0.7, 800)	MLE	14.8	0.009	94	-0.2	0.093	89	10.2	0.017	96	-0.2	0.034	98
	Profile	14.9	0.010	93	-0.1	0.113	89	10.4	0.019	98	-0.1	0.038	96
	IRLS	14.8	0.010	93	-0.2	0.102	89	10.3	0.018	97	-0.2	0.035	96
	Bayesian P1	7.7	0.003	96	0.7	0.024	98	-11.5	0.007	95	-3.1	0.052	88
	Bayesian P2	14.1	0.006	93	1.3	0.070	94	8.6	0.013	94	1.6	0.037	98
(0.3, 300,0.7, 500)	MLE	13.7	0.007	98	0.7	0.116	95	3.5	0.006	95	-0.1	0.031	96
	Profile	13.8	0.008	98	0.8	0.139	96	3.6	0.007	94	-0.1	0.034	94
	IRLS	13.8	0.007	98	0.8	0.127	96	3.6	0.006	94	-0.1	0.032	94
	Bayesian P1	10.8	0.003	97	4.1	0.078	95	-5.6	0.003	96	-0.5	0.012	99
	Bayesian P2	10.4	0.004	98	2.7	0.097	99	6.3	0.005	95	0.9	0.029	96
(0.5, 500,0.5, 500)	MLE	1.2	0.006	100	0.4	0.070	91	4.4	0.007	97	-0.1	0.061	94
	Profile	1.8	0.008	99	0.5	0.080	90	4.5	0.008	97	0.0	0.074	92
	IRLS	1.5	0.007	99	0.4	0.075	90	4.4	0.007	97	-0.1	0.067	92
	Bayesian P1	-5.5	0.003	99	-0.4	0.020	99	-2.8	0.003	99	-0.7	0.019	98
	Bayesian P2	4.3	0.004	99	2.2	0.082	91	6.8	0.005	99	1.5	0.060	94
(0.7, 800,0.5, 400)	MLE	5.5	0.014	97	-0.2	0.048	95	8.5	0.008	91	-0.1	0.059	94
	Profile	6.2	0.017	97	-0.2	0.055	93	8.4	0.008	92	-0.1	0.067	92
	IRLS	5.9	0.015	97	-0.2	0.051	93	8.4	0.008	92	-0.1	0.062	92
	Bayesian P1	-14.8	0.009	92	-3.4	0.065	88	1.1	0.003	98	0.7	0.022	99
	Bayesian P2	11.6	0.010	99	1.4	0.047	99	9.1	0.006	95	1.3	0.059	93

[a] MLE: Maximum likelihood estimate; Profile: Profile likelihood based estimate and confidence interval; IRLS: iteratively re-weighted least squares; Bayesian P1 and P2 refer to two settings of prior choice; Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$.

[b] RB: relative bias; MSE: mean squared error; CP: coverage probability.

Table A.4: Summary statistics across the five inference methods for the **two point sources adjacent category model (homogeneousness)**, based on $R = 500$ simulations with sample size $\boldsymbol{N = 1000}$.

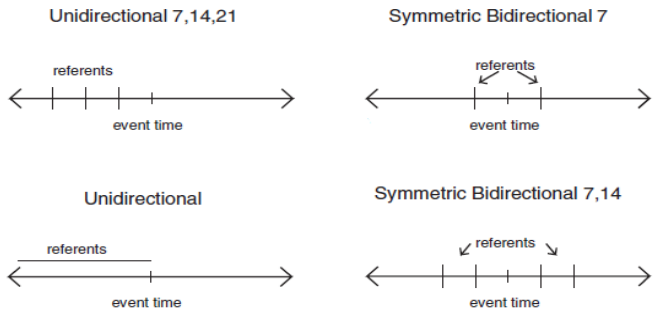
Sample size N=500		$u_{11} = log(1+\alpha_{11})$			$v_{11} = log(\beta_{11})$			$u_{12} = log(1+\alpha_{12})$			$v_{12} = log(\beta_{12})$		
$(\alpha_{11}, \beta_{11}, \alpha_{12}, \beta_{12})$	Method[a]	RB[b](%)	MSE[b]	CP[b](%)	RB(%)	MSE	CP(%)	RB(%)	MSE	CP(%)	RB(%)	MSE	CP(%)
(0.7, 800, 0.4, 300)	MLE	15.3	0.044	96	0.1	0.073	93	13.8	0.010	94	0.7	0.094	93
	Profile	15.3	0.041	95	0.0	0.075	94	12.4	0.010	95	0.6	0.091	94
	IRLS	15.6	0.039	95	0.0	0.072	94	13.2	0.009	95	0.6	0.084	94
	Bayesian P1	-18.1	0.012	91	-4.1	0.088	86	4.5	0.003	98	4.1	0.072	93
	Bayesian P2	16.6	0.015	96	2.1	0.047	98	11.9	0.006	94	2.6	0.074	97
(0.5, 500, 0.3,300)	MLE	11.6	0.009	99	0.5	0.062	94	19.1	0.011	94	-0.4	0.127	94
	Profile	12.0	0.010	98	0.4	0.061	98	17.1	0.009	95	-0.3	0.130	95
	IRLS	12.0	0.010	98	0.5	0.062	98	17.8	0.009	94	-0.4	0.130	95
	Bayesian P1	0.8	0.002	99	-0.5	0.010	100	13.4	0.003	94	4.7	0.087	97
	Bayesian P2	15.4	0.007	95	2.6	0.065	96	14.9	0.004	96	3.5	0.113	98
(0.4, 500, 0.4, 500)	MLE	15.6	0.013	94	0.8	0.199	89	19.0	0.031	98	0.2	0.157	93
	Profile	15.6	0.013	93	-0.1	0.117	91	17.6	0.018	98	-0.3	0.130	92
	IRLS	15.1	0.012	93	0.0	0.104	91	16.0	0.014	98	-0.4	0.113	93
	Bayesian P1	5.7	0.002	96	-0.7	0.017	100	6.0	0.002	99	-0.7	0.014	99
	Bayesian P2	18.3	0.007	94	3.2	0.091	96	23.1	0.009	95	4.2	0.107	99
(0.4, 500, 0.7, 800)	MLE	10.1	0.008	96	0.1	0.107	92	16.2	0.058	97	0.0	0.107	88
	Profile	10.0	0.008	95	-0.1	0.104	94	16.5	0.058	95	-0.2	0.092	93
	IRLS	10.2	0.008	95	-0.2	0.105	94	15.8	0.056	95	-0.3	0.090	93
	Bayesian P1	3.3	0.002	97	-0.7	0.016	99	-24.3	0.018	87	-4.6	0.103	81
	Bayesian P2	14.8	0.005	97	3.1	0.100	95	18.9	0.014	98	2.9	0.059	100

[a] MLE: Maximum likelihood estimate; Profile: Profile likelihood based estimate and confidence interval; IRLS: iteratively re-weighted least squares; Bayesian P1 and P2 refer to two settings of prior choice; Prior 1: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.25, 150)$; Prior 2: $(\mu_\alpha, \mu_\beta) = (0.5, 400)$ and $(\sigma^2_\alpha, \sigma^2_\beta) = (0.5, 400)$.

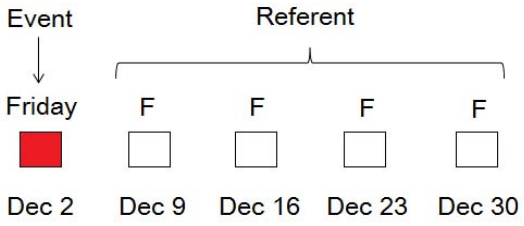
[b] RB: relative bias; MSE: mean squared error; CP: coverage probability.

A.2 Supplementary Figures and Tables for Chapter 3

Figure A.4: Referent time selections for case-crossover designs.



(a) non-localizable designs



(b) time-stratified design

Table A.5: Posterior distributions derived under full likelihood $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$ with six different prior distributions on $\boldsymbol{\nu}$ as a sensitivity analysis, under a time-stratified case-crossover design for the DAMAT study.

	$\beta_{PM_{2.5}}$			α			Number of clusters		
Priors[b]	Mean[a]	SD[a]	Median[a]	Mean	SD	Median	Mean	SD	Median
prior A	0.059	0.016	0.058				1.00	0.00	1.00
prior B	0.058	0.016	0.058				238.64	13.22	238
prior G1	0.058	0.016	0.058	0.08	0.12	0.04	1.09	0.33	1
prior G2	0.058	0.015	0.057	0.34	0.27	0.28	1.43	0.80	1
prior G3	0.058	0.016	0.058	0.08	0.11	0.04	1.06	0.27	1
prior G4	0.058	0.016	0.058	5.62	1.82	5.34	15.50	7.45	14

[a] posterior mean, standard deviation and median.

[b] Dirichlet process prior $DP(\alpha, G_0)$ on $\boldsymbol{\nu}$ in $L^T_{full}(\boldsymbol{\beta}, \boldsymbol{\nu})$. The base prior setting on G_0 was used as described in section 6; the priors on α was varied as follows.
prior A: $\nu_t = \nu^*$ for $t = 1, ..., 1096$, where $\nu^* \sim N(0, 10^2)$; prior B: $\nu_t \overset{iid}{\sim} N(0, 10^2)$;
prior G1: $\alpha \sim Gamma(0.5, 0.1)$; prior G2: $\alpha \sim Gamma(2, 0.2)$;
prior G3: $\alpha \sim Gamma(10, 0.5)$; prior G4: $\alpha \sim Gamma(20, 1)$.

A.3 Supplementary Figures and Tables for Chapter 4

Figure A.5: Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study. **Setting (a), under both assumptions 1 and 2.**

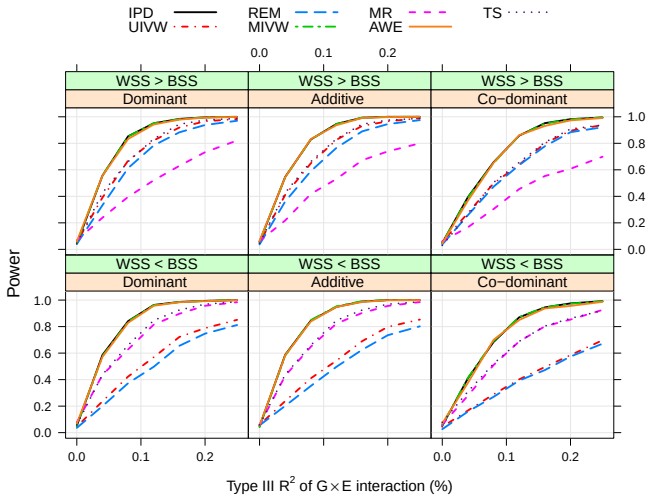


Figure A.6: Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study. **Setting (b), under assumption 1 but not 2.**

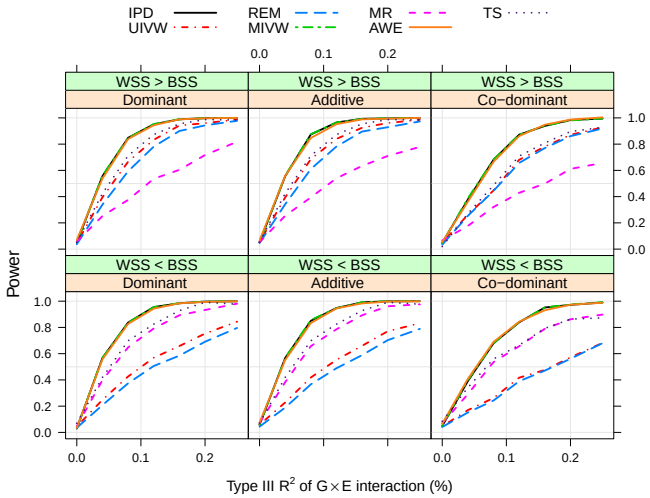


Figure A.7: Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study. **Setting (c), under assumption 2 but not 1.**

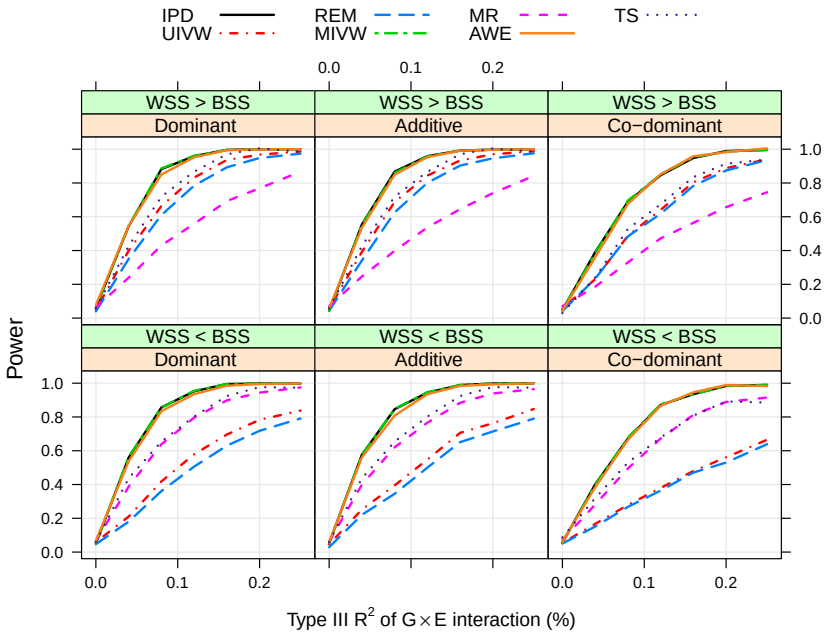


Figure A.8: Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study, for the situation of lack of common set of confounders to adjust under both assumptions 1 and 2 (**setting a**).

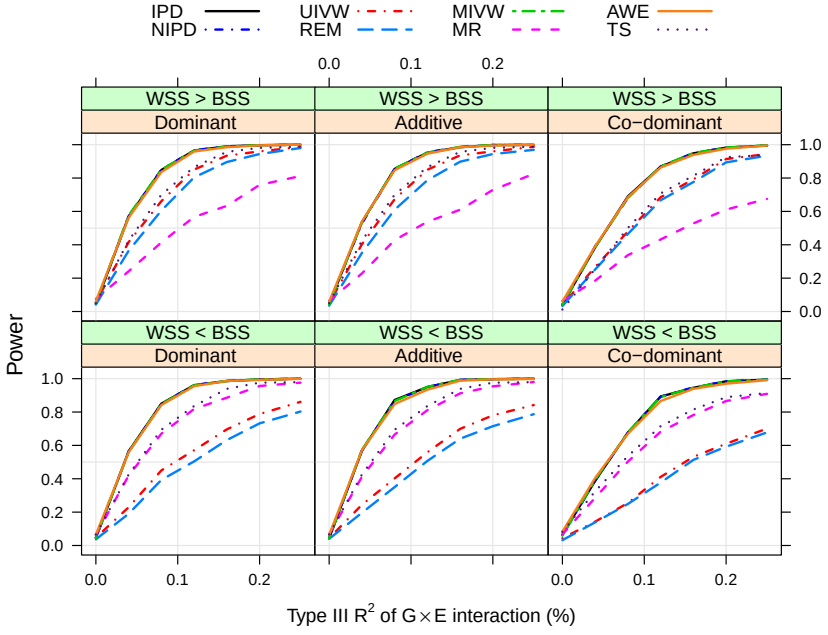


Figure A.9: Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study, for the situation of lack of common set of confounders to adjust under assumption 1 but not 2 (**setting b**).

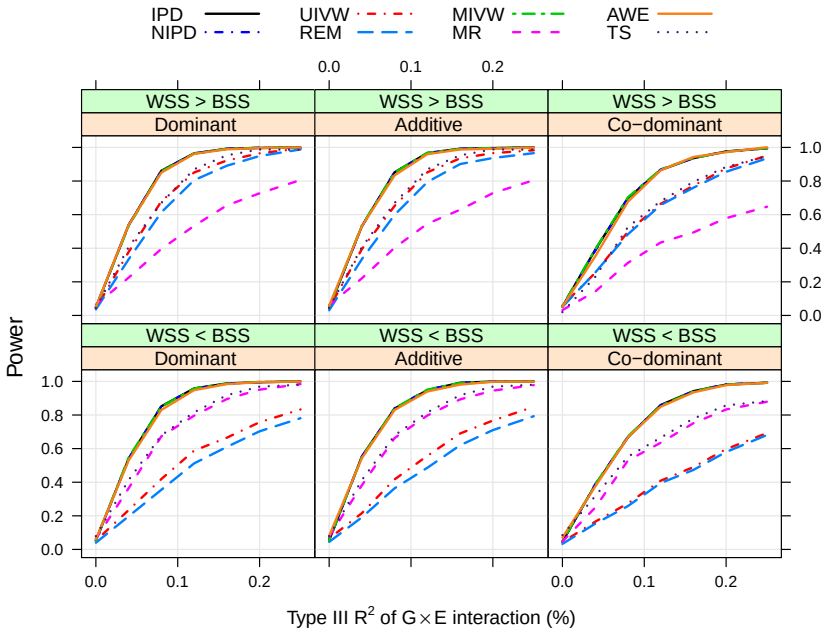


Figure A.10: Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study, for the situation of lack of common set of confounders to adjust under assumption 2 but not 1 (**setting c**).

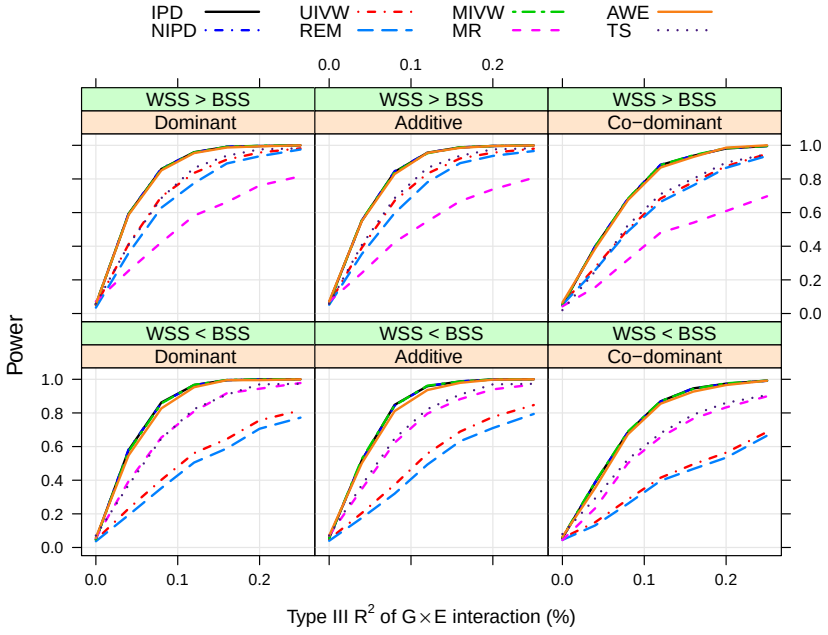


Figure A.11: Power curves under misspecified susceptibility models (dominant/additive), where the generating co-dominant model has $\delta^{AA} = 1.5\delta^{Aa}$, and no assumption of gene-environment independence or homogeneity in allele frequencies is assumed.

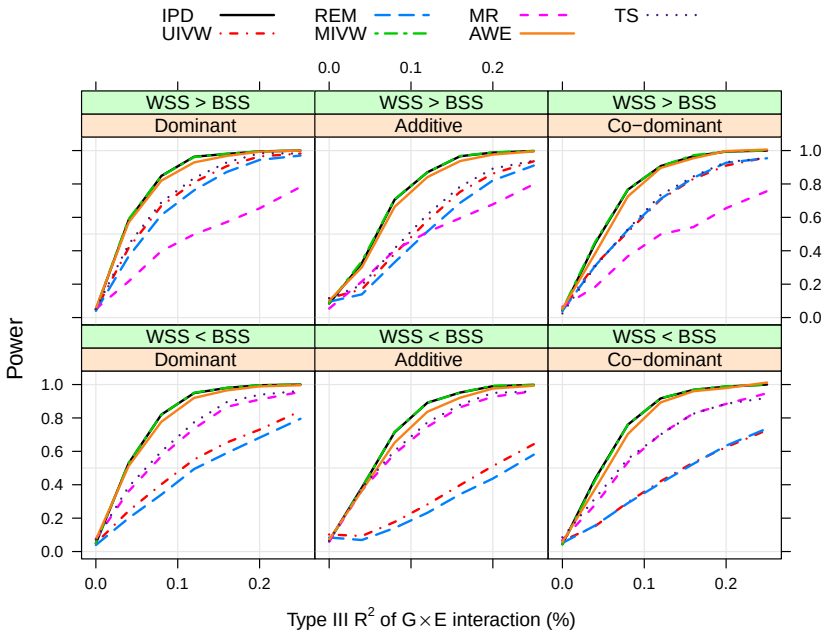


Figure A.12: Marginal SNP (rs1121980) effect against mean covariate values of age and BMI across cohorts in the FUSION study. Solid line: Meta-regression line; Dashed line: Meta-regression line without outlier (cohort FUSION-FS for age and cohort DPS for BMI).

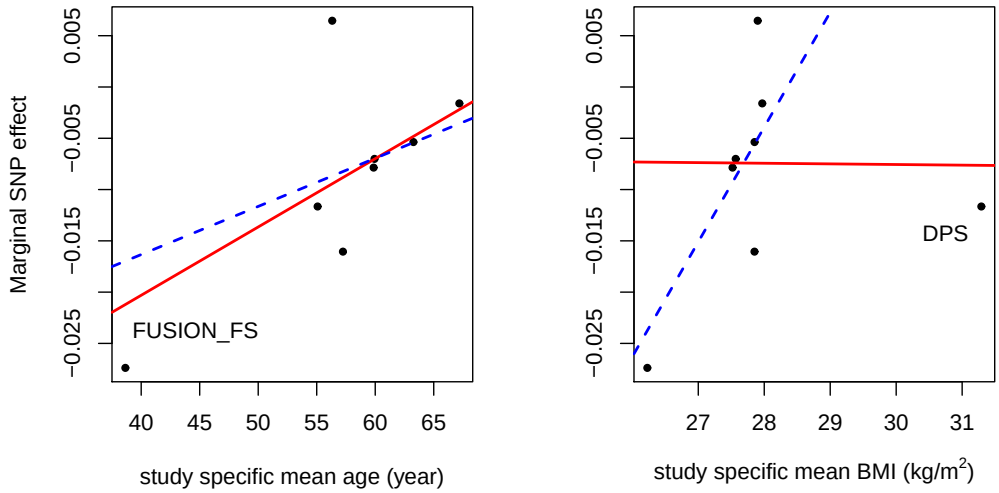


Table A.6: Comparison of the proposed meta-analytical methods under different scenarios of susceptibility models and covariate heterogeneity through a simulation study. **Setting (a): under both assumptions 1 and 2.**

	Method[a]		E(WSS/BSS)=2					E(BSS/WSS)=2				
			RB[b](%)	MV[b]	EV[b]	MSE[b]	Power	RB (%)	MV	EV	MSE	Power
R^2=0	Dominant	IPD		1.14	1.12	1.12	0.04		1.15	1.12	1.12	0.05
		UIVW		1.73	1.71	1.71	0.05		3.39	3.46	3.46	0.05
		REM		2.08	1.78	1.78	0.04		4.07	3.58	3.57	0.04
		MIVW		1.13	1.12	1.12	0.05		1.15	1.12	1.12	0.05
		MR		3.91	4.06	4.06	0.07		1.95	1.82	1.82	0.08
		AWE		1.13	1.19	1.19	0.06		1.16	1.19	1.19	0.07
	Additive	IPD		0.75	0.75	0.75	0.06		0.76	0.68	0.68	0.05
		UIVW		1.15	1.11	1.11	0.05		2.25	2.39	2.38	0.06
		REM		1.38	1.15	1.15	0.04		2.69	2.44	2.44	0.06
		MIVW		0.75	0.75	0.75	0.06		0.76	0.68	0.68	0.04
		MR		2.62	2.55	2.55	0.06		1.25	1.24	1.25	0.06
		AWE		0.75	0.79	0.79	0.06		0.76	0.72	0.72	0.05
R^2= 0.05%	Dominant	IPD	0.44	1.14	1.15	1.15	0.56	-1.78	1.13	1.29	1.29	0.59
		UIVW	1.60	1.71	1.75	1.75	0.39	0.11	3.31	3.28	3.28	0.24
		REM	1.67	2.06	1.79	1.79	0.33	0.02	3.98	3.42	3.42	0.20
		MIVW	0.42	1.13	1.15	1.15	0.56	-1.78	1.13	1.29	1.29	0.58
		MR	0.27	4.05	4.20	4.20	0.24	-2.14	1.90	2.01	2.01	0.44
		AWE	0.50	1.13	1.21	1.21	0.56	-1.90	1.13	1.36	1.36	0.57
	Additive	IPD	0.64	0.75	0.76	0.76	0.55	-0.81	0.75	0.79	0.79	0.59
		UIVW	-0.85	1.14	1.14	1.14	0.41	2.12	2.19	2.35	2.35	0.24
		REM	-0.52	1.34	1.17	1.16	0.36	2.11	2.60	2.38	2.38	0.20
		MIVW	0.71	0.75	0.76	0.76	0.54	-0.66	0.75	0.80	0.80	0.58
		MR	5.59	2.64	2.89	2.90	0.22	-1.25	1.26	1.24	1.24	0.43
		AWE	1.15	0.75	0.81	0.81	0.54	-0.59	0.75	0.83	0.83	0.58
R^2= 0.15%	Dominant	IPD	1.29	1.11	1.26	1.26	0.98	-0.28	1.14	1.44	1.44	0.99
		UIVW	1.22	1.69	1.78	1.78	0.92	-1.14	3.31	3.38	3.38	0.72
		REM	1.62	2.01	1.84	1.84	0.88	-1.50	4.00	3.52	3.53	0.66
		MIVW	1.33	1.10	1.26	1.26	0.98	-0.34	1.14	1.44	1.44	0.98
		MR	1.50	3.95	4.57	4.57	0.63	0.02	1.95	2.46	2.46	0.90
		AWE	1.47	1.11	1.34	1.34	0.98	-0.29	1.14	1.53	1.53	0.98
	Additive	IPD	-0.20	0.73	0.76	0.76	0.99	-0.32	0.75	0.95	0.95	0.99
		UIVW	0.33	1.12	1.10	1.10	0.93	0.09	2.20	2.46	2.46	0.69
		REM	0.58	1.33	1.14	1.14	0.89	0.27	2.68	2.58	2.58	0.62
		MIVW	-0.18	0.73	0.76	0.76	0.99	-0.30	0.75	0.96	0.96	0.99
		MR	-2.43	2.59	2.61	2.61	0.67	0.06	1.30	1.45	1.45	0.90
		AWE	-0.39	0.74	0.81	0.81	0.99	-0.36	0.76	1.02	1.02	0.99
R^2= 0.25%	Dominant	IPD	0.44	1.11	1.30	1.30	1.00	0.66	1.11	1.71	1.71	1.00
		UIVW	0.36	1.68	1.97	1.97	0.98	0.12	3.28	3.79	3.79	0.85
		REM	0.30	1.97	2.00	2.00	0.97	-0.00	3.93	4.02	4.02	0.81
		MIVW	0.47	1.11	1.30	1.30	1.00	0.70	1.11	1.72	1.72	1.00
		MR	0.20	4.02	4.31	4.31	0.82	0.75	1.90	2.48	2.48	0.98
		AWE	0.50	1.12	1.37	1.37	1.00	0.73	1.12	1.79	1.79	1.00
	Additive	IPD	0.13	0.74	0.87	0.86	1.00	-0.30	0.74	1.02	1.02	1.00
		UIVW	-0.34	1.12	1.22	1.22	0.99	-0.31	2.18	2.49	2.48	0.85
		REM	-0.40	1.31	1.26	1.26	0.97	-0.42	2.67	2.60	2.60	0.80
		MIVW	0.12	0.73	0.87	0.87	1.00	-0.27	0.74	1.03	1.02	1.00
		MR	0.48	2.73	2.79	2.79	0.80	0.02	1.25	1.59	1.58	0.98
		AWE	-0.01	0.75	0.89	0.89	1.00	-0.03	0.74	1.07	1.07	1.00

[a] IPD: individual patient data analysis; UIVW: univariate inverse-variance weighted estimator; REM: random effect model; MIVW: multivariate inverse-variance weighted estimator; MR: Meta-regression; AWE: adaptively weighted estimator.

[b] RB: relative bias; MV: mean of model based variance; EV: empirical variance; MSE: mean squared error. (MV, EV and MSE have been multiplied by 100.)

Table A.7: Comparison of the proposed meta-analytical methods under different scenarios of susceptibility models and covariate heterogeneity through a simulation study. **Setting (b): under assumption 1 but not 2.**

			E(WSS/BSS)=2					E(BSS/WSS)=2				
	Method[a]		RB[b](%)	MV[b]	EV[b]	MSE[b]	Power	RB (%)	MV	EV	MSE	Power
R^2= 0	Dominant	IPD		1.16	1.21	1.21	0.05		1.18	1.04	1.04	0.03
		UIVW		1.79	1.89	1.88	0.04		3.49	3.35	3.36	0.04
		REM		2.09	1.92	1.92	0.04		4.15	3.42	3.43	0.04
		MIVW		1.16	1.21	1.21	0.05		1.17	1.04	1.04	0.03
		MR		4.18	4.20	4.20	0.06		1.96	1.78	1.78	0.05
		AWE		1.19	1.28	1.28	0.07		1.18	1.08	1.08	0.04
	Additive	IPD		0.75	0.75	0.75	0.05		0.77	0.81	0.81	0.06
		UIVW		1.17	1.18	1.18	0.05		2.28	2.50	2.50	0.06
		REM		1.37	1.19	1.19	0.05		2.75	2.56	2.56	0.04
		MIVW		0.75	0.75	0.75	0.05		0.77	0.81	0.81	0.06
		MR		2.70	2.54	2.54	0.05		1.32	1.31	1.31	0.07
		AWE		0.77	0.81	0.81	0.06		0.78	0.86	0.86	0.07
R^2= 0.05	Dominant	IPD	-0.13	1.15	1.15	1.15	0.56	-1.93	1.16	1.23	1.23	0.57
		UIVW	0.88	1.78	1.86	1.86	0.40	-2.00	3.45	3.65	3.65	0.25
		REM	0.89	2.13	1.91	1.91	0.34	-1.88	4.15	3.74	3.74	0.21
		MIVW	-0.13	1.15	1.16	1.16	0.56	-1.95	1.16	1.23	1.23	0.57
		MR	-0.80	4.11	4.00	4.00	0.25	-2.55	2.00	2.02	2.02	0.40
		AWE	0.45	1.17	1.26	1.26	0.54	-2.03	1.19	1.30	1.30	0.56
	Additive	IPD	-0.78	0.75	0.79	0.79	0.56	-1.15	0.76	0.83	0.83	0.57
		UIVW	0.05	1.16	1.22	1.21	0.39	-0.10	2.27	2.40	2.39	0.23
		REM	0.06	1.39	1.28	1.28	0.34	1.20	2.72	2.51	2.51	0.19
		MIVW	-0.84	0.75	0.80	0.80	0.56	-1.24	0.76	0.83	0.83	0.57
		MR	-3.05	2.72	2.79	2.79	0.25	-1.25	1.29	1.34	1.34	0.38
		AWE	-0.98	0.77	0.86	0.86	0.56	-0.81	0.77	0.87	0.87	0.55
R^2= 0.15	Dominant	IPD	-0.00	1.14	1.16	1.16	0.99	0.82	1.15	1.51	1.51	0.98
		UIVW	-1.33	1.75	1.67	1.68	0.94	0.28	3.42	3.73	3.73	0.66
		REM	-1.15	2.10	1.73	1.73	0.90	0.41	4.09	3.81	3.81	0.58
		MIVW	-0.02	1.13	1.17	1.17	0.99	0.80	1.15	1.52	1.52	0.98
		MR	2.56	4.23	3.95	3.96	0.60	0.99	2.02	2.25	2.25	0.89
		AWE	-0.13	1.17	1.24	1.24	0.99	0.90	1.17	1.56	1.56	0.98
	Additive	IPD	-0.65	0.74	0.82	0.82	0.99	0.57	0.75	0.97	0.96	0.99
		UIVW	-0.81	1.15	1.23	1.23	0.92	1.35	2.24	2.34	2.34	0.66
		REM	-0.90	1.37	1.26	1.26	0.90	1.30	2.71	2.44	2.44	0.59
		MIVW	-0.73	0.74	0.82	0.82	0.99	0.57	0.75	0.97	0.97	0.99
		MR	-0.29	2.76	2.58	2.58	0.63	-0.05	1.32	1.51	1.51	0.89
		AWE	-0.63	0.77	0.87	0.87	0.99	0.36	0.77	1.01	1.01	0.98
R^2= 0.25	Dominant	IPD	-0.38	1.13	1.18	1.18	1.00	-0.32	1.13	1.75	1.75	1.00
		UIVW	-0.95	1.74	1.85	1.85	0.99	-0.33	3.38	4.08	4.07	0.84
		REM	-1.08	2.06	1.91	1.91	0.98	-0.48	4.07	4.17	4.16	0.80
		MIVW	-0.35	1.12	1.19	1.19	1.00	-0.39	1.13	1.76	1.76	1.00
		MR	-0.58	4.08	4.39	4.39	0.82	-0.92	1.97	2.57	2.57	0.98
		AWE	-0.59	1.15	1.28	1.28	1.00	-0.43	1.17	1.86	1.86	1.00
	Additive	IPD	0.77	0.74	0.84	0.84	1.00	1.17	0.74	1.15	1.15	1.00
		UIVW	0.07	1.15	1.24	1.24	0.99	1.71	2.22	2.69	2.70	0.83
		REM	0.16	1.37	1.26	1.26	0.97	1.55	2.67	2.72	2.72	0.79
		MIVW	0.81	0.74	0.84	0.84	1.00	1.10	0.74	1.15	1.15	1.00
		MR	2.13	2.76	2.86	2.87	0.78	0.88	1.28	1.81	1.81	0.98
		AWE	0.74	0.77	0.90	0.90	1.00	1.34	0.76	1.20	1.21	1.00

[a] IPD: individual patient data analysis; UIVW: univariate inverse-variance weighted estimator; REM: random effect model; MIVW: multivariate inverse-variance weighted estimator; MR: Meta-regression; AWE: adaptively weighted estimator.

[b] RB: relative bias; MV: mean of model based variance; EV: empirical variance; MSE: mean squared error. (MV, EV and MSE have been multiplied by 100.)

Table A.8: Comparison of the proposed meta-analytical methods under different scenarios of susceptibility models and covariate heterogeneity through a simulation study. **Setting (c): under assumption 2 but not 1.**

			E(WSS/BSS)=2					E(BSS/WSS)=2				
	Method[a]		RB[b](%)	MV[b]	EV[b]	MSE[b]	Power	RB (%)	MV	EV	MSE	Power
R^2= 0	Dominant	IPD		1.21	1.30	1.30	0.06		1.19	1.24	1.24	0.05
		UIVW		1.92	1.98	1.98	0.05		3.76	3.95	3.96	0.06
		REM		2.28	2.03	2.03	0.04		4.56	4.13	4.14	0.05
		MIVW		1.21	1.30	1.30	0.06		1.19	1.24	1.24	0.05
		MR		4.01	3.82	3.82	0.06		2.05	1.97	1.97	0.07
		AWE		1.22	1.41	1.41	0.08		1.24	1.41	1.41	0.07
	Additive	IPD		0.76	0.73	0.73	0.04		0.76	0.75	0.75	0.05
		UIVW		1.18	1.17	1.17	0.06		2.30	2.24	2.24	0.05
		REM		1.40	1.20	1.20	0.04		2.78	2.26	2.25	0.03
		MIVW		0.76	0.73	0.73	0.04		0.76	0.75	0.75	0.05
		MR		2.61	2.56	2.56	0.07		1.37	1.44	1.44	0.06
		AWE		0.76	0.79	0.79	0.06		0.80	0.83	0.83	0.06
R^2= 0.05	Dominant	IPD	2.60	1.20	1.34	1.34	0.54	1.44	1.17	1.20	1.20	0.56
		UIVW	0.14	1.90	1.99	1.98	0.40	4.08	3.73	3.74	3.74	0.21
		REM	0.09	2.28	2.05	2.05	0.35	3.14	4.46	3.84	3.84	0.18
		MIVW	2.68	1.19	1.35	1.35	0.55	1.34	1.17	1.20	1.20	0.56
		MR	6.78	3.99	4.18	4.20	0.24	0.55	2.09	2.20	2.19	0.39
		AWE	2.25	1.21	1.39	1.39	0.54	1.06	1.24	1.36	1.36	0.54
	Additive	IPD	1.66	0.75	0.78	0.78	0.55	-1.80	0.76	0.86	0.86	0.58
		UIVW	2.22	1.16	1.24	1.24	0.39	-5.26	2.28	2.30	2.30	0.25
		REM	2.64	1.38	1.25	1.25	0.34	-4.94	2.73	2.38	2.39	0.22
		MIVW	1.85	0.75	0.78	0.78	0.55	-1.73	0.75	0.86	0.86	0.57
		MR	2.33	2.67	2.83	2.83	0.25	0.72	1.36	1.40	1.40	0.40
		AWE	2.56	0.76	0.85	0.85	0.54	-1.42	0.80	0.96	0.96	0.56
R^2= 0.15	Dominant	IPD	-0.20	1.18	1.20	1.20	1.00	-0.93	1.19	1.63	1.63	0.99
		UIVW	-0.32	1.86	1.85	1.85	0.94	-1.91	3.66	4.26	4.26	0.70
		REM	-0.10	2.23	1.93	1.93	0.89	-1.68	4.42	4.35	4.35	0.63
		MIVW	-0.17	1.18	1.20	1.20	1.00	-0.95	1.18	1.62	1.62	0.99
		MR	-0.78	4.09	3.86	3.85	0.69	-0.52	2.19	2.67	2.67	0.90
		AWE	-0.11	1.21	1.26	1.26	0.99	-0.82	1.26	1.72	1.72	0.98
	Additive	IPD	0.10	0.74	0.76	0.76	0.99	-0.39	0.76	1.08	1.08	0.99
		UIVW	0.12	1.14	1.15	1.15	0.93	-1.12	2.23	2.35	2.35	0.71
		REM	-0.01	1.36	1.19	1.19	0.90	-1.13	2.72	2.48	2.48	0.65
		MIVW	0.12	0.74	0.77	0.77	0.99	-0.34	0.76	1.08	1.08	0.99
		MR	-0.29	2.69	2.65	2.65	0.64	-0.19	1.43	1.83	1.83	0.88
		AWE	-0.26	0.75	0.80	0.80	0.99	-0.34	0.81	1.19	1.19	0.98
R^2= 0.25	Dominant	IPD	-0.66	1.15	1.35	1.35	1.00	-0.70	1.13	1.77	1.77	1.00
		UIVW	-0.51	1.84	2.09	2.09	0.99	-1.03	3.60	4.24	4.24	0.84
		REM	-0.29	2.16	2.15	2.15	0.97	-1.43	4.31	4.34	4.35	0.79
		MIVW	-0.71	1.15	1.36	1.36	1.00	-0.70	1.12	1.79	1.79	1.00
		MR	-1.69	3.94	4.03	4.04	0.87	-1.01	1.99	2.61	2.62	0.97
		AWE	-1.10	1.18	1.49	1.49	1.00	-1.19	1.20	1.94	1.95	1.00
	Additive	IPD	-1.31	0.73	0.88	0.89	1.00	0.93	0.73	1.04	1.04	1.00
		UIVW	-1.19	1.13	1.27	1.27	0.99	1.15	2.21	2.49	2.49	0.85
		REM	-1.29	1.35	1.31	1.31	0.98	1.20	2.67	2.56	2.56	0.79
		MIVW	-1.36	0.72	0.89	0.89	1.00	0.90	0.72	1.04	1.04	1.00
		MR	-2.50	2.56	2.84	2.85	0.84	0.26	1.38	1.71	1.71	0.96
		AWE	-1.68	0.74	0.95	0.96	1.00	0.46	0.79	1.19	1.19	1.00

[a] IPD: individual patient data analysis; UIVW: univariate inverse-variance weighted estimator; REM: random effect model; MIVW: multivariate inverse-variance weighted estimator; MR: Meta-regression; AWE: adaptively weighted estimator.

[b] RB: relative bias; MV: mean of model based variance; EV: empirical variance; MSE: mean squared error. (MV, EV and MSE have been multiplied by 100.)

Table A.9: Comparison of the proposed meta-analytical methods under different scenarios of susceptibility models and covariate heterogeneity through a simulation study. **Setting (d): without assumption 1 or 2.**

	Method[a]		E(WSS/BSS)=2					E(BSS/WSS)=2				
			RB[b](%)	MV[b]	EV[b]	MSE[b]	Power	RB (%)	MV	EV	MSE	Power
R^2= 0	Dominant	IPD		1.23	1.21	1.21	0.04		1.22	1.16	1.16	0.04
		UIVW		1.99	2.08	2.08	0.06		3.89	4.02	4.02	0.05
		REM		2.40	2.13	2.13	0.04		4.68	4.18	4.18	0.04
		MIVW		1.22	1.22	1.22	0.04		1.22	1.16	1.16	0.04
		MR		4.05	4.13	4.13	0.07		2.12	2.05	2.06	0.06
		AWE		1.26	1.35	1.35	0.05		1.29	1.33	1.33	0.06
	Additive	IPD		0.76	0.74	0.74	0.05		0.78	0.76	0.75	0.04
		UIVW		1.20	1.21	1.21	0.05		2.35	2.42	2.42	0.06
		REM		1.43	1.24	1.24	0.04		2.84	2.48	2.48	0.04
		MIVW		0.76	0.74	0.74	0.04		0.77	0.75	0.75	0.04
		MR		2.67	2.66	2.66	0.07		1.38	1.35	1.35	0.06
		AWE		0.78	0.84	0.84	0.05		0.81	0.85	0.85	0.06
R^2= 0.05	Dominant	IPD	-2.87	1.21	1.21	1.21	0.59	0.54	1.21	1.34	1.34	0.55
		UIVW	-3.68	1.95	2.11	2.12	0.42	1.82	3.81	4.03	4.03	0.23
		REM	-3.78	2.34	2.18	2.19	0.37	1.18	4.66	4.12	4.12	0.19
		MIVW	-2.96	1.20	1.22	1.22	0.59	0.51	1.21	1.35	1.34	0.56
		MR	0.19	4.17	4.03	4.02	0.26	-0.59	2.16	2.45	2.45	0.40
		AWE	-3.18	1.25	1.32	1.32	0.57	0.04	1.29	1.46	1.46	0.53
	Additive	IPD	0.23	0.75	0.80	0.80	0.55	-0.16	0.77	0.83	0.83	0.57
		UIVW	1.14	1.18	1.22	1.22	0.37	0.85	2.31	2.46	2.46	0.23
		REM	1.40	1.40	1.27	1.27	0.33	1.21	2.80	2.52	2.52	0.19
		MIVW	0.35	0.75	0.81	0.81	0.55	-0.18	0.77	0.84	0.84	0.57
		MR	0.93	2.67	2.96	2.95	0.24	-0.10	1.44	1.35	1.35	0.37
		AWE	0.84	0.77	0.86	0.86	0.54	-0.15	0.82	0.93	0.93	0.54
R^2= 0.15	Dominant	IPD	0.52	1.19	1.25	1.25	0.99	0.09	1.17	1.51	1.51	0.98
		UIVW	1.30	1.92	2.00	2.01	0.91	0.03	3.75	4.13	4.13	0.65
		REM	1.18	2.28	2.08	2.08	0.87	0.16	4.50	4.21	4.21	0.59
		MIVW	0.58	1.18	1.26	1.26	0.99	0.14	1.17	1.52	1.52	0.98
		MR	-0.82	4.04	4.28	4.28	0.67	-0.31	2.07	2.45	2.45	0.90
		AWE	0.62	1.23	1.36	1.36	0.99	0.01	1.25	1.63	1.63	0.98
	Additive	IPD	0.85	0.74	0.78	0.78	0.99	0.74	0.75	0.86	0.86	0.99
		UIVW	0.40	1.17	1.24	1.24	0.92	1.29	2.27	2.28	2.28	0.67
		REM	0.63	1.41	1.29	1.28	0.87	1.40	2.73	2.36	2.37	0.59
		MIVW	0.85	0.74	0.78	0.78	0.99	0.73	0.75	0.86	0.86	0.99
		MR	1.62	2.70	2.56	2.56	0.65	0.05	1.40	1.60	1.60	0.87
		AWE	0.78	0.76	0.86	0.86	0.98	0.87	0.80	0.95	0.95	0.97
R^2= 0.25	Dominant	IPD	-0.54	1.18	1.36	1.36	1.00	-0.04	1.16	1.62	1.62	1.00
		UIVW	0.09	1.90	2.17	2.17	0.98	-0.13	3.75	4.32	4.32	0.82
		REM	0.15	2.23	2.19	2.19	0.97	-0.21	4.54	4.45	4.45	0.77
		MIVW	-0.54	1.18	1.37	1.37	1.00	-0.05	1.15	1.63	1.63	1.00
		MR	-1.68	4.22	4.43	4.43	0.83	0.18	2.04	2.66	2.66	0.98
		AWE	-0.63	1.23	1.51	1.51	1.00	0.19	1.24	1.82	1.82	1.00
	Additive	IPD	-1.22	0.74	0.85	0.85	1.00	-0.28	0.74	1.13	1.13	1.00
		UIVW	-0.98	1.15	1.22	1.22	0.99	-0.50	2.28	2.73	2.73	0.83
		REM	-0.98	1.38	1.27	1.27	0.97	-0.65	2.74	2.75	2.75	0.80
		MIVW	-1.26	0.73	0.85	0.85	1.00	-0.28	0.74	1.14	1.14	1.00
		MR	-2.67	2.73	3.00	3.02	0.83	0.32	1.41	1.84	1.84	0.96
		AWE	-1.23	0.76	0.93	0.93	1.00	-0.12	0.81	1.25	1.25	1.00

[a] IPD: individual patient data analysis; UIVW: univariate inverse-variance weighted estimator; REM: random effect model; MIVW: multivariate inverse-variance weighted estimator; MR: Meta-regression; AWE: adaptively weighted estimator.

[b] RB: relative bias; MV: mean of model based variance; EV: empirical variance; MSE: mean squared error. (MV, EV and MSE have been multiplied by 100.)

APPENDIX B

(Technical Details)

B.1 Technical details for chapter 2

B.1.1 Computational details for Bayesian inference

(i) Random walk Metropolis-Hastings algorithm

Without loss of generality, take one point source homogeneous ACM as an example of two-parameter model. The following mutually independent priors are considered

$$\begin{aligned} \log(1+\alpha) = u &\sim N(\mu_u, \sigma_u^2), \\ \log(\beta) = v &\sim N(\mu_v, \sigma_v^2). \end{aligned}$$

The joint posterior distribution of (u, v) is

$$\pi(u,v|X,Y) \propto \underbrace{\pi(u,v)}_{\text{Prior}} \times \underbrace{\prod_{i=1}^{N} \frac{[1+(\exp(u)-1)\cdot\exp(-(x_{i1}/\exp(v))^2)]^{k_i}}{\sum_{j=1}^{M+1}[1+(\exp(u)-1)\cdot\exp(-(x_{ij}/\exp(v))^2)]^{k_i}} \times \underbrace{\exp(Ku+Kv)}_{\text{Jacobian}}}_{\text{likelihood in terms of (u,v)}}.$$

A random walk Metropolis-Hastings algorithm is used to generate the desired draws from the target posterior distribution. The bivariate normal distribution $BVN(\boldsymbol{M}, \boldsymbol{V})$ is chosen as the proposal distribution. Initially, the mean $\boldsymbol{M^{(0)}}$ and variance $\boldsymbol{V^{(0)}}$ of this bivariate normal distribution are estimated using the Laplace transform of the posterior distribution. Then the Random walk Metropolis-Hastings algorithm can be proceeded as follows:

- Step 1: The Markov chain starts with a random initial guess of $(u^{(0)}, v^{(0)})$.

- Step 2: In the *t*-th step ($t \geq 1$), the proposal distribution follows $(u^{(t-1)}, v^{(t-1)})^T + \omega \boldsymbol{Z}$, where ω is the step size and $\boldsymbol{Z} \sim BVN(\boldsymbol{0}, \boldsymbol{V})$. A random draw (u^*, v^*) is generated from this proposal distribution.

- Step 3: The ratio $R = \pi(u^*, v^*|X, Y)/\pi(u^{(t-1)}, v^{(t-1)}|X, Y)$ is calculated, and the draw (u^*, v^*) is accepted with probability of $P = min(R, 1)$, i.e,

$$\begin{aligned} (u^{(t)}, v^{(t)}) &= (u^*, v^*), && \text{if} \quad 0 \leq \pi \leq P, \\ &= (u^{(t-1)}, v^{(t-1)}), && \text{if} \quad P < \pi \leq 1, \end{aligned}$$

  where $\pi \sim Uniform(0, 1)$.

- Step 4: Step 2 and 3 was repeated T times to generate a Markov chain of length T, i.e. $(u^{(t)}, v^{(t)})$, $t = 1, 2, ..., T$.

The accept ratio of (u, v) is defined as the proportion of times that the proposed draws (u^*, v^*)'s are accepted. The step size ω in step 2 is chosen such that the accept ratios laid within $(0.25, 0.40)$.

(ii) Metropolis-Hastings within Gibbs algorithm

Without loss of generality, take one point source ACM as an illustration example. The following mutually independent priors are considered

$$\begin{aligned} \log(1 + \alpha_i) = u_i &\sim N(\mu_{u_i}, \sigma^2_{u_i}), \\ \log(\beta_i) = v_i &\sim N(\mu_{v_i}, \sigma^2_{v_i}), \quad i = 1, ..., K. \end{aligned}$$

The joint posterior distribution of $(\boldsymbol{u}, \boldsymbol{v})$ is

$$\pi(\boldsymbol{u}, \boldsymbol{v}|X, Y) \propto \underbrace{\pi(\boldsymbol{u}, \boldsymbol{v})}_{\text{Prior}} \times \underbrace{\exp \sum_{i=1}^{K} (u_i + v_i)}_{\text{Jacobian}} \times \underbrace{\prod_{i=1}^{N} \frac{\prod_{h=1}^{k_i} [1 + (\exp(u_h) - 1) \cdot \exp(-(x_{i1}/\exp(v_h))^2)]}{\sum_{j=1}^{M+1} \prod_{h=1}^{k_i} [1 + (\exp(u_h) - 1) \cdot \exp(-(x_{ij}/\exp(v_h))^2)]}}_{\text{likelihood in terms of (u,v)}}.$$

A Metropolis-Hastings within Gibbs algorithm is used to generate the draws from the above posterior distribution. In the *t*-th step of the block Gibbs algorithm, $(u_1^{(t)}, v_1^{(t)})$ is supposed to be drawn from the conditional distribution $\pi(u_1, v_1|u_2^{(t-1)}, v_2^{(t-1)}, ..., u_K^{(t-1)}, v_K^{(t-1)}, X, Y)$, and then $(u_2^{(t)}, v_2^{(t)})$ is supposed to be drawn from the conditional distribution $\pi(u_2, v_2|u_1^{(t)}, v_1^{(t)}, u_3^{(t-1)}, v_3^{(t-1)}, ..., u_K^{(t-1)}, v_K^{(t-1)}, X, Y)$ with updated draws of $(u_1^{(t)}, v_1^{(t)})$, and so on. However, the above conditional densities are not from standard distributions. Instead, random walk Metropolis-Hastings algorithm is used to generate the desired draws from these full conditional distribution. The bivariate normal distribution $BVN(\boldsymbol{M}, \boldsymbol{V})$ is chosen as the proposal distribution. Initially, the mean $\boldsymbol{M^{(0)}}$ and variance $\boldsymbol{V^{(0)}}$ of this bivariate normal distribution are estimated using the Laplace transform of the posterior distribution. The Metropolis-Hastings within Gibbs algorithm can be proceeded as follows:

- Step 1: The Markov chain starts with a random initial guess of $(\boldsymbol{u}^{(0)}, \boldsymbol{v}^{(0)})$.
- Step 2: In the *t*-th step, $(\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)})$ is supposed to be generated by

  drawing $(u_1^{(t)}, v_1^{(t)})$ from $\pi(u_1, v_1|u_2^{(t-1)}, v_2^{(t-1)}, ..., u_K^{(t-1)}, v_K^{(t-1)}, X, Y)$

  drawing $(u_2^{(t)}, v_2^{(t)})$ from $\pi(u_2, v_2|u_1^{(t)}, v_1^{(t)}, u_3^{(t-1)}, v_3^{(t-1)}, ..., u_K^{(t-1)}, v_K^{(t-1)}, X, Y)$

  ...

  drawing $(u_K^{(t)}, v_K^{(t)})$ from $\pi(u_K, v_K|u_1^{(t)}, v_1^{(t)}, ..., u_{K-1}^{(t)}, v_{K-1}^{(t)}, X, Y)$

However, these conditional densities are not from standard distributions. Step 2 is actually proceeded as

- Step 2.1: The proposal distribution of the full conditional $\pi(u_1, v_1 | u_2^{(t-1)}, v_2^{(t-1)}, ..., u_K^{(t-1)}, v_K^{(t-1)}, X, Y)$ follows $(u_1^{(t-1)}, v_1^{(t-1)})^T + \omega_1 \boldsymbol{Z}$, where ω_1 is the step size for (u_1, v_1) and $\boldsymbol{Z} \sim BVN(\boldsymbol{0}, \boldsymbol{V})$. A random draw (u_1^*, v_1^*) is generated from this proposal distribution.
- Step 2.2: The ratio $R = \frac{\pi(u_1^*, v_1^* | u_2^{(t-1)}, v_2^{(t-1)}, ..., u_K^{(t-1)}, v_K^{(t-1)}, X, Y)}{\pi(u_1^{(t-1)}, v_1^{(t-1)} | u_2^{(t-1)}, v_2^{(t-1)}, ..., u_K^{(t-1)}, v_K^{(t-1)}, X, Y)}$ is calculated.
- Step 2.3: (u_1^*, v_1^*) is accepted with probability of $P = min(R, 1)$, i.e,

$$\begin{aligned} (u_1^{(t)}, v_1^{(t)}) &= (u_1^*, v_1^*), && \text{if} \quad 0 \leq \pi \leq P, \\ &= (u_1^{(t-1)}, v_1^{(t-1)}), && \text{if} \quad P < \pi \leq 1, \end{aligned}$$

  where $\pi \sim Uniform(0, 1)$.
- Step 2.4: Step 2.1-2.3 is repeated for the other pairs of sub-parameter (u_2, v_2) to (u_K, v_K) to generate $(\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)}) = (u_1^{(t)}, v_1^{(t)}, ..., u_K^{(t)}, v_K^{(t)})$.

- Step 3: Step 2 is repeated T times to generate a Markov chain of length T, i.e. $(\boldsymbol{u}^{(t)}, \boldsymbol{v}^{(t)})$, $t = 1, 2, ..., T$.

The accept ratio of (u_k, v_k) is defined as the proportion of times that the proposed draws (u_k^*, v_k^*) is accepted, k=1,...,K. The step size ω_k in step 2 is chosen such that the accept ratios laid within $(0.25, 0.40)$.

B.2 Technical details for chapter 3

B.2.1 Proofs of the equivalence results

Frequentist equivalence using full likelihood

$\log(L^T_{full}(\boldsymbol{\beta},\nu)) = \sum_{t=1}^{T} Y_t\Big[\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t - \sum_{s\in W(t)} \log\{1+\exp(\nu_t+\boldsymbol{\beta}^\top \boldsymbol{X}_s)\}\Big]$ and then the estimating equation for $\boldsymbol{\beta}$ is

$$\begin{aligned}
U^T_{full}(\boldsymbol{\beta}) &= \sum_{t=1}^{T} Y_t \boldsymbol{X}_t - \sum_{t=1}^{T}\sum_{s=1}^{T} Y_t \frac{I(s\in W(t))\exp(\nu_t+\boldsymbol{\beta}^\top \boldsymbol{X}_s)\boldsymbol{X}_s}{1+\exp(\nu_t+\boldsymbol{\beta}^\top \boldsymbol{X}_s)} \\
&= \sum_{t=1}^{T} Y_t \boldsymbol{X}_t - \sum_{t=1}^{T}\sum_{s=1}^{T} Y_s \frac{I(t\in W(s))\exp(\nu_s+\boldsymbol{\beta}^\top \boldsymbol{X}_t)\boldsymbol{X}_t}{1+\exp(\nu_s+\boldsymbol{\beta}^\top \boldsymbol{X}_t)} \\
&= \sum_{t=1}^{T} Y_t \boldsymbol{X}_t - \sum_{t=1}^{T} \boldsymbol{X}_t \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\Big\{\sum_{s=1}^{T} \frac{Y_s I(s\in R(t))\exp(\nu_s)}{1+\exp(\nu_s+\boldsymbol{\beta}^\top \boldsymbol{X}_t)}\Big\} \\
&= \sum_{t=1}^{T} \boldsymbol{X}_t \Big\{Y_t - \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\sum_{s\in R(t)} \frac{Y_s\exp(\nu_s)}{1+\exp(\nu_s+\boldsymbol{\beta}^\top \boldsymbol{X}_t)}\Big\},
\end{aligned}$$

where $R(t)$ is the set of days that contain day t in their reference window. For SBD and TSD but not more generally, $R(t) = W(t)$ (Lu and Zeger 2007). Comparing with the log-linear model estimating equation corresponding to β

$$U_{ll}(\boldsymbol{\beta}) = \sum_{t=1}^{T} \boldsymbol{X}_t \Big\{Y_t - \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S_t)\Big\},$$

So, if $\widehat{S}_t(\nu,\boldsymbol{\beta}) = \log(\sum_{s\in R(t)} Y_s\exp(\nu_s)/\{1+\exp(\nu_s+\boldsymbol{\beta}^\top \boldsymbol{X}_t)\})$, then $U^T_{full}(\boldsymbol{\beta})$ will provide the same estimate for $\boldsymbol{\beta}$ as $U_{ll}(\boldsymbol{\beta})$. Under a TSD, while the conditional likelihood approach or an equivalent log-linear model would only allow the risk changes abruptly among different time stratifications, the full likelihood approach does not require such constraint because $\widehat{S}_{t'}(\boldsymbol{\nu},\boldsymbol{\beta})$ is not necessarily equal to $\widehat{S}_t(\boldsymbol{\nu},\boldsymbol{\beta})$ for $t'\in W(t)$ and $t'\neq t$.

Bayesian equivalence using conditional likelihood

Proof of Theorem 3.1:

$$L_{ll}(\boldsymbol{\beta}, S_t) \;\propto\; \prod_{k=1}^{K} \prod_{t:\, t\in ts(k)} \big\{ \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S'_k)\big\}^{Y_t} \exp\{-\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S'_k)\}.$$

Let $\varphi_k = \exp(S'_k)$. The marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L_{ll}(\boldsymbol{\beta}, S_t)$ is

$$\begin{aligned}
& \pi(\boldsymbol{\beta} \mid \boldsymbol{X}, \boldsymbol{Y}) \propto \int \pi(\boldsymbol{\beta})\pi(S'_1, ..., S'_K) L_{ll}(\boldsymbol{\beta}, S_t) dS'_1 \cdots dS'_K \\
\propto\;& \pi(\boldsymbol{\beta}) \int \prod_{k=1}^{K} \prod_{t:\, t\in ts(k)} \big\{ \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S'_k)\big\}^{Y_t} \exp\{-\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t + S'_k)\} dS'_1 \cdots dS'_K \\
=\;& \pi(\boldsymbol{\beta}) \big[\prod_{t=1}^{T} \big\{ \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\big\}^{Y_t} \big] \prod_{k=1}^{K} \int \varphi_k^{-1} \prod_{t:\, t\in ts(k)} \big[\varphi_k^{Y_t} \exp\{-\varphi_k \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\}\big] d\varphi_k \\
=\;& \pi(\boldsymbol{\beta}) \big[\prod_{t=1}^{T} \big\{ \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\big\}^{Y_t} \big] \prod_{k=1}^{K} \int \big[\varphi_k^{\sum_{t:\, t\in ts(k)} Y_t - 1} \exp\{-\varphi_k \sum_{t:\, t\in ts(k)} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\}\big] d\varphi_k \\
=\;& \pi(\boldsymbol{\beta}) \big[\prod_{t=1}^{T} \big\{ \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\big\}^{Y_t} \big] \prod_{k=1}^{K} \{ \sum_{t:\, t\in ts(k)} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\}^{-\sum_{t\in ts(k)} Y_t} \\
=\;& \pi(\boldsymbol{\beta}) \big[\prod_{t=1}^{T} \big\{ \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\big\}^{Y_t} \big] \prod_{k=1}^{K} \prod_{t:\, t\in ts(k)} \{ \sum_{s\in W(t)} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_s)\}^{-Y_t} \\
=\;& \pi(\boldsymbol{\beta}) \big[\prod_{t=1}^{T} \big\{ \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)\big\}^{Y_t} \big] \prod_{t=1}^{T} \{ \sum_{s\in W(t)} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_s)\}^{-Y_t} \\
=\;& \pi(\boldsymbol{\beta}) \prod_{t=1}^{T} \Big\{ \frac{\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_t)}{\sum_{s\in W(t)} \exp(\boldsymbol{\beta}^\top \boldsymbol{X}_s)} \Big\}^{Y_t} \\
=\;& \pi(\boldsymbol{\beta}) L_{cc}(\boldsymbol{\beta}),
\end{aligned}$$

which is the marginal posterior distribution of $\boldsymbol{\beta}$ derived from $L_{cc}(\boldsymbol{\beta})$.

Bayesian equivalence using full likelihood

Let $y_{s\cdot t} = y_{s1t} + y_{s0t}$, and let $\Phi_{st} = \exp(\phi_{st})$. Then

$$\begin{aligned}
&\pi(\boldsymbol{\nu}, \boldsymbol{\beta} \mid \boldsymbol{X}, \boldsymbol{Y}) \propto \int \pi(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta}) L_p(\boldsymbol{\phi}, \boldsymbol{\nu}, \boldsymbol{\beta}) d\boldsymbol{\phi} \propto \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) \prod_{t=1}^{T} \big\{ \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t) \big\}^{Y_t} \\
&\times \int \pi(\boldsymbol{\phi}) \prod_{s=1}^{T} \prod_{t=1}^{T} [\exp(\phi_{st})]^{y_{s\cdot t}} \exp\{-\exp(\phi_{st})[1 + \exp(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t)]\} d\boldsymbol{\phi} \\
= \;& \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) \prod_{t=1}^{T} \big\{ \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t) \big\}^{Y_t} \prod_{s=1}^{T} \prod_{t=1}^{T} \int \Phi_{st}^{y_{s\cdot t}-1} \exp\big\{ -\Phi_{st}[1 + \exp(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t)] \big\} d\Phi_{st} \\
= \;& \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) \prod_{t=1}^{T} \big\{ \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t) \big\}^{Y_t} \prod_{s=1}^{T} \prod_{t=1}^{T} \Big\{ \frac{1}{1 + \exp(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t)} \Big\}^{y_{s\cdot t}} \\
= \;& \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) \prod_{t=1}^{T} \big\{ \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t) \big\}^{Y_t} \prod_{s=1}^{T} \prod_{t=1}^{T} \Big\{ \frac{1}{1 + \exp(\nu_s + \boldsymbol{\beta}^\top \boldsymbol{X}_t)} \Big\}^{y_s I(s \in R(t))} \\
= \;& \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) \prod_{t=1}^{T} \big\{ \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t) \big\}^{Y_t} \prod_{t=1}^{T} \prod_{s=1}^{T} \Big\{ \frac{1}{1 + \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_s)} \Big\}^{y_t I(s \in W(t))} \\
= \;& \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) \prod_{t=1}^{T} \Big[\big\{ \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t) \big\}^{Y_t} \prod_{s \in W(t)} \Big\{ \frac{1}{1 + \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_s)} \Big\}^{y_t} \Big] \\
= \;& \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) \prod_{t=1}^{T} \Big[\frac{\exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_t)}{\prod_{s \in W(t)} \{1 + \exp(\nu_t + \boldsymbol{\beta}^\top \boldsymbol{X}_s)\}} \Big]^{Y_t} \\
= \;& \pi(\boldsymbol{\nu}, \boldsymbol{\beta}) L_{full}^{T}(\boldsymbol{\nu}, \boldsymbol{\beta})
\end{aligned}$$

B.2.2 Computational details for Bayesian inference

Metropolis-Hastings within Gibbs algorithm

Sampling of $\boldsymbol{\beta}$ under conditional likelihood formulation: We take conditional likelihood under the shared exposure as an example. For mutually independent normal priors $\boldsymbol{\beta} \sim N(\boldsymbol{\mu}_\beta, \sigma_\beta^2 \boldsymbol{I}_p)$, the joint posterior distribution of $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^\top$ is not a standard distribution. Let $\pi(\boldsymbol{\theta} \mid \cdot)$ denote the full conditional distribution as a function of $\boldsymbol{\theta}$ given the data and all other parameters. The posterior distribution $\pi(\boldsymbol{\beta} \mid \boldsymbol{X}, \boldsymbol{Y})$ can be obtained

using a Gibbs sampler through the following full conditional distributions,

$$\pi(\beta_r|\cdot) = \exp\Big\{-\frac{(\beta_r-\mu_{\beta_r})^2}{2\sigma^2_{\beta_r}}\Big\}\prod_{t=1}^{T}\Big\{\frac{\exp(\beta_r X_{tr})}{\sum_{s\in W(t)}\exp(\boldsymbol{\beta}^\top \boldsymbol{X}_s)}\Big\}^{Y_t}, \quad r=1,...,p.$$

To sample from $\pi(\beta_r \mid \cdot)$, we followed these steps of a Metropolis-Hastings algorithm:

- Step 1. Start with initial value $\boldsymbol{\beta}^{(0)} = (\beta_1^{(0)},...,\beta_p^{(0)})^\top$.
- Step 2. For $r = 1,...,p$, at the k-th iteration with the current value as $\boldsymbol{\beta}^{(k)} = (\beta_1^{(k)},...,\beta_{r-1}^{(k)},\beta_r^{(k-1)},...,\beta_p^{(k-1)})^\top$. Generate a new value β_r^* from a candidate density $g(\beta_r)$ and replace $\beta_r^{(k)}$ by β_r^* with probability $\min\Big\{1, \frac{\pi(\beta_r^*|\cdot)g(\beta_r^{(k-1)})}{\pi(\beta_r^{(k-1)}|\cdot)g(\beta_r^*)}\Big\}$. We chose the candidate density $g(\beta_r)$ as the prior density $\pi(\beta_r)$. Since the full conditional density $\pi(\beta_r \mid \cdot) \propto \pi(\beta_r)L(\beta_r \mid \cdot)$, then the acceptance probability reduces to $\min\Big\{1, \frac{L(\beta_r^*|\cdot)}{L(\beta_r^{(k-1)}|\cdot)}\Big\}$. Then $\beta_r^{(k)}, 1, ..., p$, is updated accordingly.
- Step 3. Run the chain with 10,000 iterations.

Sampling of $(\boldsymbol{\nu},\boldsymbol{\beta})$ under full likelihood formulation: We take full likelihood $L^T_{full}(\boldsymbol{\beta},\boldsymbol{\nu})$ under the shared exposure as an example. To sample from the posterior distribution of $\boldsymbol{\nu}$ and $\boldsymbol{\beta}$, we adopted a componentwise Metropolis-Hastings algorithm. The full conditional distributions used are:

$$\pi(\beta_r|\cdot) \propto \exp\Big\{-\frac{(\beta_r-\mu_{\beta_r})^2}{2\sigma^2_{\beta_r}}\Big\}\prod_{t=1}^{T}\Big\{\frac{\exp(\beta_r X_{tr})}{\prod_{s\in W(t)}[1+\exp(\nu_t+\boldsymbol{\beta}^\top \boldsymbol{X}_s)]}\Big\}^{y_t}, \; r=1,....,p$$

$$\pi(\nu_t|\cdot) \propto \Big\{\frac{1}{\prod_{s\in W(t)}[1+\exp(\nu_t+\boldsymbol{\beta}^\top \boldsymbol{X}_s)]}\Big\}^{y_t} \times$$

$$\Big\{\frac{\alpha}{T-1+\alpha}\frac{\exp(\mu y_t+\frac{y_t^2\sigma^2}{2})}{\sqrt{2\pi\sigma^2}}\times\exp\Big[-\frac{(\nu_t-\mu-y_t\sigma^2)^2}{2\sigma^2}\Big]+\frac{\exp(y_t\nu_t)}{T-1+\alpha}\sum_{s=1,\,s\neq t}^{T} I(\nu_s=\nu_t)\Big\},$$

At each iteration, we first update the value of $\boldsymbol{\beta}$ similarly as described above, and then move on to the cycle for $\boldsymbol{\nu}$ with the updated value of $\boldsymbol{\beta}$ substituted. Particularly, given current values of $\boldsymbol{\beta}$, $\boldsymbol{\nu}$ is updated in the following way:

- Step 1. As a metropolis-Hastings step, ν_t^* was drawn from the candidate distribution of $\pi(\nu_t \mid \boldsymbol{\nu}_{-t})$, namely from $\frac{\alpha}{T-1+\alpha}N(\mu,\sigma^2) + \frac{1}{T-1+\alpha}\sum_{s=1,\, s\neq t}^{T} I(\nu_s = \nu_t)$. In particular, one either get a distinct value for ν_t^* from the normal component with probability $\frac{\alpha}{T-1+\alpha}$ or get a draw of ν_t^* with equal probability from the current set of the $(T-1)$ entries of $\boldsymbol{\nu}_{-t}$. We adopt the algorithm in to generate observations from this candidate density.

- Step 2. Set the new value of ν_t to ν_t^*, with acceptance probability $\min\left\{1, \frac{L(\nu_t^*|\cdot)}{L(\nu_t^{'}|\cdot)}\right\}$, where $\nu_t^{'}$ is the current working value of ν_t.

- Step 3. Repeat steps 1-2 5 times and consider the last of these updates of ν_t as the value of $\nu_t^{(k)}$, say at the k-th iteration.

- Step 4. Repeat steps 1-3 for all $\nu_t^{(k)}$ for $t = 1, ..., T$. One complete iteration of the Markov chain consists of the foregoing updates for both the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$. Given current values of $\boldsymbol{\beta}^{(k)}$ and $\boldsymbol{\nu}^{(k)}$, we can go to the next iteration for $\boldsymbol{\beta}^{(k+1)}$ and $\boldsymbol{\nu}^{(k+1)}$. We run the chain with 10,000 iterations.

Prior choices under the simulation study

Assume the informative prior on β has the form $\beta \sim N(\mu_\beta, \sigma_\beta^2)$. According to the ad-hoc prior eliciting strategy for the DAMAT study, when $\beta^* = 0.1$ we a priori postulated a 95% confidence interval (1.02, 1.15) for $\exp(\beta)$, and solved for the approximated values of $(\mu_\beta, \sigma_\beta)$ as $(0.05, 0.02)$. Thus our informative prior was chosen as $N(0.08, 0.03^2)$ when $\beta^* = 0.1$. Similarly, when $\beta^* = 1$ we presumed a 95% confidence interval $(0.4, 1.2)$ for $\exp(\beta)$ and deduced the corresponding informative prior $\beta \sim N(0.8, 0.2^2)$. To complete the hierarchy, we have used $\alpha \sim Gamma(2, 0.1)$; $G_0 \sim N(\mu, \sigma^2)$, $\mu \sim N(0, 10)$ and $\sigma^{-2} \sim Gamma(4, 1)$ in all our simulations.

Construction of power priors for the DAMAT study data example

Asthma risk has been associated with $PM_{2.5}$/PM_{10} in many studies using both time-series and case-crossover designs. Among recent papers, an Alaskan study (Chimonas et al., 2007) found that a 10 $\mu g\, m^{-3}$ increase in PM_{10} was associated with a 0.6% (95% CI: 0.1%, 1.3%) increase in outpatient asthma visits, and a 1.8% (95% CI: 0.6%, 3.0%) increase in inhaled quick-relief mediation prescriptions. In Rio de Janeiro (Moura et al., 2009), a 10 $\mu g\, m^{-3}$ increase of PM_{10} was found to be associated with 6.7% (95% CI: 1.8%, 11.5%) increase for bronchial obstruction. In two Idaho cities (Ulirsch et al., 2007), a 24.3 $\mu g\, m^{-3}$ increase in PM_{10} was associated with 4.3% increase for respiratory disease. In the Detroit Medicaid population (Li et al., 2011), we found a 3-7% increase in asthma risks for a 9.2 $\mu g\, m^{-3}$ increase in $PM_{2.5}$. Larger effects were found when only the warmer season was considered (Villeneuve et al., 2007). These results are converted in terms of risk ratios in the following table. More detailed reviews can be found in Li et al. (2011).

Study	Risk Ratios*
Chimonas et al., 2007	1.006, 1.018
Moura et al., 2009	1.065
Ulirsch et al., 2007	1.017
Li et al., 2011	1.03-1.09

* Risk ratios $\exp(\hat{\beta}_{PM_{2.5}})$ and $\exp(\hat{\beta}_{PM_{10}})$ corresponding to 10 $\mu g\, m^{-3}$ increase in PM_x concentrations

Based on these studies where different cohorts, statistical models, and variant asthma outcomes were used, we have a belief that the asthma-$PM_{2.5}$ association is in general modest with an odds ratio ranging (1.01-1.09) for a 10 $\mu g\, m^{-3}$ increase in $PM_{2.5}$ (if we assume that effect of PM_{10} has no substantial difference from that of $PM_{2.5}$). In our DAMAT data analysis section, we constructed a presumed 95% confidence interval (1.01,1.09) based on the above information, and took the prior mean to be the center

($\mu_\beta = [\log(1.09) + \log(1.01)]/2 = 0.05$) and the prior standard deviation to be one-fourth the width of the interval ($\sigma_\beta = [\log(1.09) - \log(1.01)]/4 = 0.02$), i.e., $\beta_{PM_{2.5}} \sim N(0.05, 0.02^2)$.

For the power priors, suppose we observed D_0 in terms of summary statistics from previous studies, e.g. the MLEs $\hat{\beta}_k$'s with variance $\hat{\sigma}^2_{\hat{\beta}_k}$'s. We assume the sampling distribution of $\hat{\beta}_k$ is normal, namely, $\hat{\beta}_k|\beta \sim N(\beta, \hat{\sigma}^2_{\hat{\beta}_k})$, $k = 1, ..., K$. Assuming the studies were independent and equally weighted, then

$$L(\beta|D_0) \propto \prod_{k=1}^{K} \exp(-\frac{(\beta - \hat{\beta}_k)^2}{2\hat{\sigma}^2_{\hat{\beta}_k}}).$$

In particular, we considered $K = 3$ prior studies having small, modest and strong effect sizes $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3) = (0.02, 0.05, 0.08)$ with $(\hat{\sigma}^2_{\hat{\beta}_1}, \hat{\sigma}^2_{\hat{\beta}_2}, \hat{\sigma}^2_{\hat{\beta}_3}) = (0.02, 0.02, 0.03)$ respectively, to reflect $PM_{2.5}$-asthma association (change in asthma risk for a 10 $\mu g\, m^{-3}$ increase in $PM_{2.5}$) based on the evidence in a recent review paper (Li et al. 2011). $L(\beta|D_0)$ is then described by the product of the three independent normal likelihoods.

B.3 Technical details for chapter 4

B.3.1 Proofs of the theoretical results

Proof of Lemma 4.1 and Theorem 4.1

(1) Proof of Lemma 4.1: Let $\boldsymbol{X}_1 = (X_1, ..., X_p)$ and $\boldsymbol{X}_2 = (X_{p+1}, ..., X_{p+q})$. Then $\hat{\boldsymbol{\zeta}} = (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{Y}$

$$\text{and} \quad \begin{pmatrix} \hat{\boldsymbol{\theta}}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1^\top \boldsymbol{X}_1 & \boldsymbol{X}_1^\top \boldsymbol{X}_2 \\ \boldsymbol{X}_2^\top \boldsymbol{X}_1 & \boldsymbol{X}_2^\top \boldsymbol{X}_2 \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{X}_1^\top \boldsymbol{Y} \\ \boldsymbol{X}_2^\top \boldsymbol{Y} \end{pmatrix}.$$

Let $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top$, then $\hat{\boldsymbol{\theta}}_2$ can be written as $(\boldsymbol{X}_2^\top \boldsymbol{A}\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{A}\boldsymbol{Y}$. We have

$$\begin{aligned} Cov(\hat{\boldsymbol{\zeta}}, \hat{\boldsymbol{\theta}}_2) &= Cov\{(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{Y}, \quad (\boldsymbol{X}_2^\top \boldsymbol{A}\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top \boldsymbol{A}\boldsymbol{Y}\} \\ &= (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top Cov(\boldsymbol{Y}, \boldsymbol{Y})\boldsymbol{A}^\top \boldsymbol{X}_2(\boldsymbol{X}_2^\top \boldsymbol{A}\boldsymbol{X}_2)^{-1} \\ &= (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top Cov(\boldsymbol{Y}, \boldsymbol{Y})\{\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\}\boldsymbol{X}_2(\boldsymbol{X}_2^\top \boldsymbol{A}\boldsymbol{X}_2)^{-1} \\ &= Var(Y_i)\{(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top - (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top\}\boldsymbol{X}_2(\boldsymbol{X}_2^\top \boldsymbol{A}\boldsymbol{X}_2)^{-1} \\ &= 0 \end{aligned}$$

Because the MLEs $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{\theta}}_2$ are asymptotically normal, then $\hat{\boldsymbol{\zeta}}$ and $\hat{\boldsymbol{\theta}}_2$ are asymptotically independent.

(2) Proof of Theorem 4.1: Follow lemma 4.1, $\hat{\delta}_k$ and $\hat{\lambda}_k$ are asymptotically independent because they come from two nested linear regression models. So $cov(\hat{\delta}_k, \hat{\lambda}_k) = 0$. We also have $cov(\hat{\delta}_j, \hat{\lambda}_k) = 0$ for $j \neq k$ among the K independent studies. Under the standard condition, $\hat{\delta}^{\text{UIVW}}$ is a linear combination of $\hat{\delta}_k$ and $\hat{\delta}^{\text{MR}}$ is a linear combination of $\hat{\lambda}_k$, it follows $cov(\hat{\delta}^{\text{UIVW}}, \hat{\delta}^{\text{MR}}) = 0$.

For $\hat{\delta}^{\text{AWE}}(w) = w\hat{\delta}^{\text{UIVW}} + (1-w)\hat{\delta}^{\text{MR}}$, $0 \leq w \leq 1$, we have $\mathbf{v}(\hat{\delta}^{\text{AWE}}(w)) = w^2\mathbf{v}(\hat{\delta}^{\text{UIVW}}) + (1-w)^2\mathbf{v}(\hat{\delta}^{\text{MR}}) + 2w(1-w)cov(\hat{\delta}^{\text{UIVW}}, \hat{\delta}^{\text{MR}}) = w^2\mathbf{v}(\hat{\delta}^{\text{UIVW}}) + (1-$

$w)^2\mathbf{v}(\hat{\delta}^{\mathrm{MR}}) = \{\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})+\mathbf{v}(\hat{\delta}^{\mathrm{MR}})\}[w-\mathbf{v}(\hat{\delta}^{\mathrm{MR}})/\{\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})+\mathbf{v}(\hat{\delta}^{\mathrm{MR}})\}]^2+\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})\mathbf{v}(\hat{\delta}^{\mathrm{MR}})$ $/\{\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})+\mathbf{v}(\hat{\delta}^{\mathrm{MR}})\}$. So $\mathbf{v}(\hat{\delta}^{\mathrm{AWE}}(w))$ reaches its minimum if and only if $w = \mathbf{v}(\hat{\delta}^{\mathrm{MR}})/$ $\{\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})\}$. With this choice of w, $\mathbf{v}(\hat{\delta}^{\mathrm{AWE}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}$.

Proof of Proposition 4.1

(1) Proof of Proposition 4.1: Assumption 4.1 implies the distributions $P(E|G = g)$ are the same for $g = 0, 1, 2$ (corresponding to (aa, Aa, AA) respectively), within each study as well as for the whole population. Let $n_{gk} = \sum_{i:G_{ki}=g} 1$, $\overline{G}_k = \sum_{i=1}^{n_k} G_{ki}/n_k$, $m_{gk} = \sum_{i:G_{ki}=g} E_{ki}/n_{gk}$; and $\mu_{gk} = E(E_{ki}|G_{ki} = g, \text{study} = k)$, $\mu_k = E(E_{ki}|\text{study} = k)$. Assumption 4.1 implies: (i) $\mu_{gk} = \mu_k$; (ii) $\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)E_{ki} \to 0$ as $n_k \to \infty$; and (iii)

$$\frac{\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)G_{ki}E_{ki}}{\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)G_{ki}} = \frac{\sum_{g=1,2}\{\sum_{i:\,G_{ki}=g}(g - \overline{G}_k)gE_{ki}\}}{\sum_{g=1,2}\{\sum_{i:\,G_{ki}=g}(g - \overline{G}_k)g\}} = \frac{\sum_{g=1,2}\{(g - \overline{G}_k)gn_{gk}m_{gk}\}}{\sum_{g=1,2}\{(g - \overline{G}_k)gn_{gk}\}} \to \mu_k,$$

as $n_{gk} \to \infty$. We have $\hat{\lambda}_k = \sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)Y_{ki}/\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)G_{ki}$. So

$$\begin{aligned} E(\hat{\lambda}_k) &= \alpha_k\frac{\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)}{\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)G_{ki}} + \beta_G + \beta_E\frac{\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)E_{ki}}{\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)G_{ki}} + \delta\frac{\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)G_{ki}E_{ki}}{\sum_{i=1}^{n_k}(G_{ki} - \overline{G}_k)G_{ki}} \\ &\to \beta_G + \delta\mu_k, \quad \text{as } n_{gk} \to \infty \end{aligned}$$

So $\hat{\delta}^{\mathrm{MR}}$ is asymptotically unbiased for δ under assumption 4.1. The above calculation holds for dominant, recessive and additive genetic susceptibility models. For co-dominant model, the calculation holds for AA and Aa respectively.

(2) Unbiased MR when assumption 4.1 is relaxed: If assumption 4.1 is relaxed, unbiased estimator of δ can still be found through different MR models for different susceptibility model. For example, under dominant model,

$$E(\hat{\lambda}_k) = \beta_G + \beta_E(\frac{n_{1k}m_{1k} + n_{2k}m_{2k}}{n_{1k} + n_{2k}} - m_{0k}) + \delta(\frac{n_{1k}m_{1k} + n_{2k}m_{2k}}{n_{1k} + n_{2k}});$$

under additive model,

$$E(\hat{\lambda}_k) = \beta_G + \beta_E\frac{-\overline{G}_k n_{0k}m_{0k} + (1 - \overline{G}_k)n_{1k}m_{1k} + (2 - \overline{G}_k)n_{2k}m_{2k}}{(1 - \overline{G}_k)n_{1k} + 2(2 - \overline{G}_k)n_{2k}} + \delta\frac{(1 - \overline{G}_k)n_{1k}m_{1k} + 2(2 - \overline{G}_k)n_{2k}m_{2k}}{(1 - \overline{G}_k)n_{1k} + 2(2 - \overline{G}_k)n_{2k}}.$$

(3) Bias of $\hat{\delta}^{\mathrm{MR}}$ in terms of bss/tss: For simplicity, we derive the bias of $\hat{\delta}^{\mathrm{MR}}$ under assumption 4.2 for a dominant model. Let the sample mean of E for the carrier (non-carrier) group be m_{1k} (m_{0k}) for the k-th study. We have

$$E(\hat{\delta}^{\mathrm{MR}}) \;=\; \{\sum_k w_k(m_k-\overline{m})^2\}^{-1}\{\sum_k w_k(m_k-\overline{m})E(\hat{\lambda}_k)\},$$

where $w_k = n_k^{-1}n_{1k}n_{0k}$, $\overline{m} = (\sum_k n_k^{-1}n_{1k}n_{0k}m_k)/(\sum_k n_k^{-1}n_{1k}n_{0k})$ and $E(\hat{\lambda}_k) = \beta_G + \beta_E(m_{1k}-m_{0k}) + \delta m_{1k}$. We have $n_{1k}/n_k \to p$, $m_k \xrightarrow{p} \mu_k$, $\overline{m} \xrightarrow{p} \mu$, $s^2_{Ek} \xrightarrow{p} \sigma^2_{Ek}$, as $n_k \to \infty$. Suppose $n_k/N \to r_k \in (0,1)$ as $N \to \infty$. Then

$$E(\hat{\delta}^{\mathrm{MR}}) \xrightarrow{p} \{\sum_k r_k(\mu_k-\mu)^2\}^{-1}\sum_k r_k(\mu_k-\mu)\{\beta_E(\mu_{1k}-\mu_{0k})+\delta\mu_{1k}\}.$$

Denote the sample Pearson correlation coefficient between G and E in study k as

$$\hat{\rho}_k = \frac{\sum_{i=1}^{n_k}(G_{ki}-\overline{G}_k)E_{ki}}{\{\sum_{i=1}^{n_k}(G_{ki}-\overline{G}_k)G_{ki}\}^{\frac{1}{2}}\{n_k s^2_{Ek}\}^{\frac{1}{2}}},$$

then we can write

$$m_{1k}-m_k \;=\; n_k^{-1}n_{0k}(m_{1k}-m_{0k}) = n_k^{-1}n_{0k}\frac{\sum_{i=1}^{n_k}(G_{ki}-\overline{G}_k)E_{ki}}{\sum_{i=1}^{n_k}(G_{ki}-\overline{G}_k)G_{ki}} = n_k^{-1}n_{0k}\hat{\rho}_k\{\frac{n_k s^2_{Ek}}{n_k^{-1}n_{0k}n_{1k}}\}^{\frac{1}{2}}.$$

Asymptotically, $\mu_{1k}-\mu_k = (1-p)(\mu_{1k}-\mu_{0k}) = (1-p)\rho_k[\sigma^2_{Ek}/\{p(1-p)\}]^{\frac{1}{2}}$. The bias of $\hat{\delta}^{\mathrm{MR}}$

$$\begin{aligned} E(\hat{\delta}^{\mathrm{MR}})-\delta \;&\xrightarrow{p}\; \{\sum_k r_k(\mu_k-\mu)^2\}^{-1}\sum_k r_k(\mu_k-\mu)\{\beta_E(\mu_{1k}-\mu_{0k})+\delta(\mu_{1k}-\mu_k+\mu)\} \\ &=\; \{p(1-p)\}^{-\frac{1}{2}}\{\sum_k r_k(\mu_k-\mu)^2\}^{-1}\sum_k r_k(\mu_k-\mu)\sigma_{Ek}[\beta_E\rho_k+\delta(1-p)\rho_k]. \end{aligned}$$

Clearly, if assumption 4.1 holds (implying $\rho_k = 0$, for $k = 1, ..., K$), $E(\hat{\delta}^{\mathrm{MR}})-\delta \xrightarrow{p} 0$. If not, we have $0 \le \beta_E\rho_k + \delta(1-p)\rho_k \le \beta_E + \delta(1-p)$. From Cauchy-Schwarz inequality, we have $\sum_k r_k(\mu_k-\mu)\sigma_{Ek} \le (bss \times wss)^{\frac{1}{2}}$, and then $\{\sum_k r_k(\mu_k-\mu)\sigma_{Ek}\}/\{\sum_k r_k(\mu_k-\mu)^2\} \le (wss/bss)^{\frac{1}{2}}$. When N is large, the limiting value of $E(\hat{\delta}^{\mathrm{MR}})-\delta$ is bounded from above by $\{\beta_E+\delta(1-p)\}\{p(1-p)\}^{-\frac{1}{2}}(wss/bss)^{\frac{1}{2}}$. Given p, β_E and δ, the upper bound

increases as wss/bss increases, or equivalently, as bss/tss decreases. So $\hat{\delta}^{\mathrm{AWE}}$ can control for the bias by putting less weight on $\hat{\delta}^{\mathrm{MR}}$ when the bias of $\hat{\delta}^{\mathrm{MR}}$ increases.

Derivation of $\hat{\mathbf{v}}(\hat{\delta})$ and $\mathbf{v}(\hat{\delta})$ under G-E independence assumption

Under the dominant model, for the k-th study, denote n_{1k} (n_{0k}) as the number of carriers (non-carrier), for $k = 1, ..., K$; denote the sample mean of E for the carrier (non-carrier) group as m_{1k} (m_{0k}), and denote the sample variance of E for carrier (non-carrier) group as s^2_{E1k} (s^2_{E0k}), where $\frac{1}{n_{1k}} \sum_{i:\, G_{ki}=1} (E_{ki} - m_{1k})^2$. Approximately, $m_{1k} = m_{0k} = m_k, s^2_{E1k} = s^2_{E0k} = s^2_{Ek}$ under assumption 4.1. $\hat{\mathbf{v}}(\hat{\delta})$ can be derived as follows:

(1) IPD analysis: Under model (1), $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{IPD}})$ is the sub 3×3 matrix of $(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\hat{\sigma}^2$, where

$$
(\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \left(\begin{array}{cccc|ccc}
n_1 & 0 & \cdots & 0 & n_{11} & n_1 m_1 & n_{11} m_1 \\
 & n_2 & & \vdots & n_{12} & n_2 m_2 & n_{12} m_2 \\
 & & \ddots & 0 & \vdots & \vdots & \vdots \\
 & & & n_K & n_{1K} & n_K m_K & n_{1K} m_K \\
\hline
 & & & & \sum_k n_{1k} & \sum_k n_{1k} m_k & \sum_k n_{1k} m_k \\
 & & & & & \sum_{k,\, i} E^2_{ki} & \sum_{k,\, i:G_{ki}=1} E^2_{ki} \\
 & & & & & & \sum_{k,\, i:G_{ki}=1} E^2_{ki}
\end{array}\right)^{-1}_{(K+3)(K+3)}
$$

$$
= \left(\begin{array}{ccc}
\ddots & & \cdots \\
\vdots & \left(\begin{array}{ccc}
\sum_k n_k^{-1} n_{0k} n_{1k} & 0 & \sum_k n_k^{-1} n_{1k} n_{0k} m_k \\
 & \sum_k n_k s^2_{Ek} & \sum_k n_{1k} s^2_{E1k} \\
 & & \sum_k (n_{1k} s^2_{E1k} + n_k^{-1} n_{1k} n_{0k} m_k^2)
\end{array}\right)^{-1}_{3\times 3} & \\
 & &
\end{array}\right)_{(K+3)(K+3)}.
$$

$$
\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}}) = \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{IPD}})_{(3,3)} = \{(\sum_k n_{1k} s^2_{E1k})(\sum_k n_{0k} s^2_{E0k})/(\sum_k n_k s^2_{Ek}) + \sum_k n_k^{-1} n_{1k} n_{0k} (m_k - \overline{m})^2\}^{-1} \hat{\sigma}^2.
$$

(2) Inverse-variance weighted estimator: $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)$ is the sub 3×3 matrix of $(X_k^\top X_k)^{-1}\hat{\sigma}_k^2$,

where $(X_k^\top X_k)^{-1} =$

$$\left(\begin{array}{c|ccc} n_k & n_{1k} & n_k m_k & n_{1k} m_k \\ \hline & n_{1k} & n_{1k} m_k & n_{1k} m_k \\ & & \sum_i E_{ki}^2 & \sum_{i:G_{ki}=1} E_{ki}^2 \\ & & & \sum_{i:G_{ki}=1} E_{ki}^2 \end{array}\right)^{-1} = \left(\begin{array}{cc} \cdot & \cdots \\ \vdots & \left(\begin{array}{ccc} n_k^{-1} n_{0k} n_{1k} & 0 & n_k^{-1} n_{1k} n_{0k} m_k \\ & n_k s_{Ek}^2 & n_{1k} s_{E1k}^2 \\ & & n_{1k} s_{E1k}^2 + n_k^{-1} n_{1k} n_{0k} m_k^2 \end{array}\right)^{-1} \end{array}\right).$$

Compare $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1}$ with $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{IPD}})^{-1}$, we have $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}}) = \{\sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1}\}^{-1} = \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{IPD}})$, which implies $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MIVW}}) = \hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}}) = \{\sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1}\}^{-1}_{(3,3)}$. For $\hat{\delta}_k$, we have $\hat{\mathbf{v}}(\hat{\delta}_k) = \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)_{(3,3)} = n_k s_{Ek}^2 \hat{\sigma}_k^2 / (n_{1k} s_{E1k}^2 n_{0k} s_{E0k}^2)$. Then

$$\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}}) = \{\sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1}\}^{-1} = \{\sum_k (n_{1k} s_{E1k}^2 n_{0k} s_{E0k}^2)/(n_k s_{Ek}^2 \hat{\sigma}_k^2)\}^{-1}.$$

(3) Meta-regression: We have $\hat{\lambda}_k = \sum_{i=1}^{n_k} (G_{ki} - \overline{G}_k) Y_{ki} / \sum_{i=1}^{n_k} (G_{ki} - \overline{G}_k) G_{ki}$ and $\hat{\mathbf{v}}(\hat{\lambda}_k) = n_{1k}^{-1} n_{0k}^{-1} n_k \hat{\sigma}_{\eta k}^2$. $\overline{m}$ can be simplified as $(\sum_k n_k^{-1} n_{1k} n_{0k} \hat{\sigma}_{\eta k}^{-2} m_k)/(\sum_k n_k^{-1} n_{1k} n_{0k} \hat{\sigma}_{\eta k}^{-2})$. Then the weighted simple linear meta-regression yields that

$$\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}}) = \{\sum_k w_k (m_k - \overline{m})^2\}^{-1} = \{\sum_k n_k^{-1} n_{1k} n_{0k} \hat{\sigma}_{\eta k}^{-2} (m_k - \overline{m})^2\}^{-1}.$$

(4) Asymptotic model based variance $\mathbf{v}(\hat{\delta})$: Suppose $n_k/N \to r_k \in (0,1)$ as $N \to \infty$. If we assume $\sigma_k^2 = \sigma^2$ for $k = 1, ..., K$, we have $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, $\hat{\sigma}_k^2 \xrightarrow{p} \sigma^2$, as $N \to \infty$. Under the IPD model, where the homoscedasticity assumption has been implicitly made for the regression model, we have $\mathbf{v}(Y_{ki}|\boldsymbol{X}_{ki}) = \sigma^2$ that does not depend on $\boldsymbol{X}_{ki}$. Therefore, $\mathbf{v}(\hat{\lambda}_k) = n_{1k}^{-1} n_{0k}^{-1} n_k \mathbf{v}(Y_{ki}) = n_{1k}^{-1} n_{0k}^{-1} n_k \sigma^2$. Moreover, under assumption 4.1, we have the facts that: (i) $m_{1k} \xrightarrow{p} \mu_k$, $m_{0k} \xrightarrow{p} \mu_k$, $m_k \xrightarrow{p} \mu_k$, as $n_k \to \infty$; (ii) $s_{E1k}^2 \xrightarrow{p} \sigma_{Ek}^2$, $s_{E0k}^2 \xrightarrow{p} \sigma_{Ek}^2$, $s_{Ek}^2 \xrightarrow{p} \sigma_{Ek}^2$, as $n_k \to \infty$. For dominant model, $n_{1k}/n_k \to p_k$, as $n_k \to \infty$. Let $\overline{\mu} = \{\sum_k n_k p_k (1-p_k) \mu_k\}/\{\sum_k n_k p_k (1-p_k)\}$. $\overline{m} \xrightarrow{p} \overline{\mu}$ as $N \to \infty$. The asymptotic

model based variance can be derived as follows

$$\begin{aligned}
\mathbf{v}(\hat{\delta}^{\mathrm{IPD}}) &= \{(\sum_k n_k p_k \sigma_{Ek}^2)(\sum_k n_k(1-p_k)\sigma_{Ek}^2)/(\sum_k n_k \sigma_{Ek}^2) + \sum_k n_k p_k(1-p_k)(\mu_k - \overline{\mu})^2\}^{-1}\sigma^2, \\
\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) &= \Big\{\sum_k n_k p_k(1-p_k)\sigma_{Ek}^2\Big\}^{-1}\sigma^2, \\
\mathbf{v}(\hat{\delta}^{\mathrm{MR}}) &= \{\sum_k n_k p_k(1-p_k)(\mu_k - \overline{\mu})^2\}^{-1}\,\sigma^2.
\end{aligned}$$

By Slutsky's theorem, for large N, $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}})$, $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MIVW}})$, $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}})$, $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})$ are consistent estimators of $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})$, $\mathbf{v}(\hat{\delta}^{\mathrm{MIVW}})$, $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})$, $\mathbf{v}(\hat{\delta}^{\mathrm{MR}})$ respectively.

Proof of Theorem 4.2

We first prove the following four propositions.

Propositions. Under assumption 4.1, we have

P1. $\mathbf{v}(\hat{\delta}^{\mathrm{MIVW}'}) = \mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'})$.

P2. $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}$.

P3. $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) \geq \mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'})$. The equality holds if and only if $p_k = p$, for $k = 1, 2, ..., K$.

P4. For $\hat{\delta}^{\mathrm{AWE2}'} = w\hat{\delta}^{\mathrm{MIVW2}'} + (1-w)\hat{\delta}^{\mathrm{MR}}$, $0 \leq w \leq 1$, we have that $\mathbf{v}(\hat{\delta}^{\mathrm{AWE2}'})^{-1}$ attains its maximum at $\mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}$ if and only if $w = \mathbf{v}(\hat{\delta}^{\mathrm{MR}})/\{\mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'}) + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})\}$.

Proof of P1: For the centered model, we have

$$(\boldsymbol{X}_k^{'\top}\boldsymbol{X}_k^{'})^{-1} \approx \begin{pmatrix} n_k & n_{1k} & 0 & 0 \\ & n_{1k} & 0 & 0 \\ & & \sum_i E_{ki}^{'2} & \sum_{i:G_{ki}=1} E_{ki}^{'2} \\ & & & \sum_{i:G_{ki}=1} E_{ki}^{'2} \end{pmatrix}^{-1}, \quad \text{then}$$

$$\mathbf{v}(\hat{\boldsymbol{\beta}}^{\mathrm{MIVW2}'}) = \big(\sum_k \mathbf{v}(\hat{\beta}_{Ek}', \hat{\delta}_k')^{-1}\big)^{-1} = \begin{pmatrix} \sum_k n_k \sigma_{Ek}^2 & \sum_k n_k p_k \sigma_{E1k}^2 \\ & \sum_k n_k p_k \sigma_{E1k}^2 \end{pmatrix}^{-1} \sigma^2 \quad \text{and}$$

$$\mathbf{v}(\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}'}) = \big(\sum_k \mathbf{v}(\hat{\boldsymbol{\beta}}'_k)^{-1}\big)^{-1} = \begin{pmatrix} \sum_k n_k p_k(1-p_k) & 0 & 0 \\ & \sum_k n_k \sigma^2_{Ek} & \sum_k n_k p_k \sigma^2_{E1k} \\ & & \sum_k n_k p_k \sigma^2_{E1k} \end{pmatrix}^{-1} \sigma^2, \text{ so}$$

$$\mathbf{v}(\hat{\delta}^{\mathrm{MIVW}'}) = \mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'}) = \frac{\sum_k n_k \sigma^2_{Ek}}{\{\sum_k n_k p_k \sigma^2_{Ek}\}\{\sum_k n_k (1-p_k)\sigma^2_{Ek}\}}\,\sigma^2.$$

There is no efficiency gain by pooling all three parameters $(\beta'_{Gk}, \beta'_E, \delta')$ over polling the two common parameters (β'_E, δ') under a centered model.

Proof of P2:

$$\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} = \Big[\underbrace{(\sum_k n_k p_k \sigma^2_{Ek})(\sum_k n_k(1-p_k)\sigma^2_{Ek})/(\sum_k n_k \sigma^2_{Ek})}_{\mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'})^{-1}\sigma^2} + \underbrace{\sum_k n_k p_k (1-p_k)(\mu_k - \overline{\mu})^2}_{\mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}\sigma^2}\Big]\sigma^{-2}.$$

Proof of P3: To show $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) \geq \mathbf{v}(\hat{\delta}^{\mathrm{MIVW2}'})$, it is sufficient to show $\big(\sum_k n_k p_k (1-p_k)\sigma^2_{Ek}\big)(\sum_k n_k \sigma^2_{Ek}) \leq (\sum_k n_k p_k \sigma^2_{Ek})(\sum_k n_k(1-p_k)\sigma^2_{Ek})$, or equivalently, $\sum_k n_k^2 p_k(1-p_k)\sigma^4_{Ek} + \sum_{i\neq j} n_i n_j p_i(1-p_i)\sigma^2_{Ei}\sigma^2_{Ej} \leq \sum_k n_k^2 p_k(1-p_k)\sigma^4_{Ek} + \sum_{i\neq j} n_i n_j p_i(1-p_j)\sigma^2_{Ei}\sigma^2_{Ej}$. Then it is sufficient to show for $\forall\, i,j,\ n_i n_j p_i(1-p_i)\sigma^2_{Ei}\sigma^2_{Ej} + n_j n_i p_j(1-p_j)\sigma^2_{Ej}\sigma^2_{Ei} \leq n_i n_j p_i(1-p_j)\sigma^2_{Ei}\sigma^2_{Ej} + n_j n_i p_j(1-p_i)\sigma^2_{Ej}\sigma^2_{Ei} \Leftrightarrow \forall\, i,j,\ n_i n_j (p_i - p_j)^2 \sigma^2_{Ei}\sigma^2_{Ej} \geq 0$, which is apparently true. All the above equalities hold if and only if $p_i = p_j,\ \forall\, i,j$, i.e., $p_k = p$, for $k = 1, 2, ..., K$.

Proof of P4: Following Lemma 4.1, $cov(\hat{\delta}_k, \hat{\lambda}_k) = 0$ and $cov(\hat{\beta}_{Ek}, \hat{\lambda}_k) = 0$. Because $\hat{\delta}^{\mathrm{MR}}$ is a linear combination of $\hat{\lambda}_k$ and $\hat{\beta}^{\mathrm{MIVW2}'}$ is a linear combination of $\hat{\beta}_{Ek}$ and $\hat{\delta}_k$, we have $cov(\hat{\delta}^{\mathrm{MIVW2}'}, \hat{\delta}^{\mathrm{MR}}) = 0$. The rest of the proof is similar as Theorem 4.1.

Denote $\hat{\delta}^{\mathrm{MIVW2}}$ as the MIVW pooling (β_E, δ) from the un-centered model. We have $\hat{\delta}^{\mathrm{MIVW2}} = \hat{\delta}^{\mathrm{MIVW2}'}$ since $(\beta_E, \delta) = (\beta'_E, \delta')$. Propositions P1-P4 still hold if $\hat{\delta}^{\mathrm{MIVW2}'}$ is substituted by $\hat{\delta}^{\mathrm{MIVW2}}$.

Proof of Theorem 4.2: Following propositions P1 and P2, $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} \geq \mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}$ under assumption 4.1. The equality holds if and only if $p_k = p$, for $k =$

$1, 2, ..., K$.

Proof of the results in Remark 4.4: According to propositions P1 and P3, $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{AWE2'}})^{-1}$ under assumption 4.1. Together with the results in Theorem 4.1 and Proposition P2, $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{AWE2'}})^{-1} \geq \mathbf{v}(\hat{\delta}^{\mathrm{AWE}})^{-1}$. The equality holds if and only if $p_k = p$, for $k = 1, 2, ..., K$.

Proof of Theorem 4.3

Proof of Theorem 4.3: Under both assumptions 1 and 2, the variance of $\hat{\delta}^{\mathrm{IPD}}$, $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$ can be further simplified as $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}}) = [p(1-p)\sum_k n_k\{\sigma_{Ek}^2 + (\mu_k - \mu)^2\}]^{-1}\sigma^2$, $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) = \{p(1-p)\sum_k n_k\sigma_{Ek}^2\}^{-1}\sigma^2$ and $\mathbf{v}(\hat{\delta}^{\mathrm{MR}}) = \{p(1-p)\sum_k n_k(\mu_k - \mu)^2\}^{-1}\sigma^2$. Now we have $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}$. From Theorem 4.1, $\mathbf{v}(\hat{\delta}^{\mathrm{AWE}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{UIVW}})^{-1} + \mathbf{v}(\hat{\delta}^{\mathrm{MR}})^{-1}$, then $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}})^{-1} = \mathbf{v}(\hat{\delta}^{\mathrm{AWE}})^{-1}$.

Suppose $n_k/N \to r_k \in (0,1)$ as $N \to \infty$. We have $m_k \xrightarrow{p} \mu_k$, $s_{Ek}^2 \xrightarrow{p} \sigma_{Ek}^2$, $m \xrightarrow{p} \mu$, $s_E^2 \xrightarrow{p} \sigma_E^2$, as $N \to \infty$. Then $TSS/N = s_E^2 \xrightarrow{p} \sigma_E^2$, $WSS/N = \sum_k n_k s_{Ek}^2/N \xrightarrow{p} \sum_k r_k\sigma_{Ek}^2$, $BSS/N = \sum_k n_k(m_k - m)^2/N \xrightarrow{p} \sum_k r_k(\mu_k - \mu)^2$, as $N \to \infty$. Because $TSS = WSS + BSS$, TSS/N is both consistent estimator of σ_E^2 and $\sum_k r_k\sigma_{Ek}^2 + \sum_k r_k(\mu_k - \mu)^2$. So we have $\sigma_E^2 = \sum_k r_k\{\sigma_{Ek}^2 + (\mu_k - \mu)^2\}$, i.e., $tss = wss + bss$. Therefore, $\mathbf{v}(\hat{\delta}^{\mathrm{UIVW}}) = \{Np(1-p)wss\}^{-1}\sigma^2$, $\mathbf{v}(\hat{\delta}^{\mathrm{MR}}) = \{Np(1-p)bss\}^{-1}\sigma^2$ and $\mathbf{v}(\hat{\delta}^{\mathrm{IPD}}) = \mathbf{v}(\hat{\delta}^{\mathrm{AWE}}) = \{Np(1-p)tss\}^{-1}\sigma^2$.

B.3.2 Details of the simulation study

For G, without assumption 4.2, the MAF q_k is generated from $U(0.15, 0.35)$ independently for each study k, and then G_{ki} is generated as (AA, Aa, aa) with probability $\left(q_k^2, 2q_k(1-q_k), (1-q_k)^2\right)$ that follows HWE. With assumption 4.2, the MAF q is generated from $U(0.15, 0.35)$, and then G_{ki} is generated as (AA, Aa, aa) with probability

$\left(q^2, 2q(1-q), (1-q)^2\right)$. Susceptibility models including dominant, additive and co-dominant models are considered under each simulation.

For E, the k-th study mean of E was sampled from $\mu_k \sim N(\mu, \sigma_\mu^2)$ with known μ and between study variance σ_μ^2. The k-th study variance $\sigma_{E_k}^2$ was sampled from $\sigma_{E_k}^2 \sim \sigma_\mu^2 \times U(c_1, c_2)$ with choices of constant (c_1, c_2) satisfy the two cases that $E(WSS/BSS) = 2$ and $E(BSS/WSS) = 2$ respectively. With assumption 4.1, the values of E of the k-th study were sampled from $E_{ki}|\mu_k, \sigma_{E_k}^2 \sim N(\mu_k, \sigma_{E_k}^2)$ that is independent of G. Without assumption 4.1, potential $G - E$ dependence was considered through the group mean μ_{gk} as follows. $(\mu_{0k}, \mu_{1k}, \mu_{2k})$ was calculated from the following equations $\mu_k n_k = \sum_{g=0,1,2} n_{gk}\mu_{gk}$ and $\mu_{2k} = \mu_{1k} + d_1\sigma_{E_k} = \mu_{0k} + d_2\sigma_{E_k}$, where $d_1 \sim U(0, 0.5)$ and $d_2 \sim U(d_1, 1)$. In general, $\sigma_{E_k} > \sigma_{E_k,\mathrm{w}} \gg \sigma_{E_k,\mathrm{w}}/\sqrt{n_{gk}}$, where the common within group variance $\sigma_{E_k,\mathrm{w}}^2$ is calculated from $n_k\sigma_{E_k}^2 = n_k\sigma_{E_k,\mathrm{w}}^2 + \sum_{g=0,1,2} n_{gk}(\mu_{gk} - \mu_k)^2$. Thus, the k-th study mean is μ_k, and the group means μ_{gk} are potentially dependent on G. Then E_{ki} were sampled from $E_{ki}|G_{ki} = g, \mu_{gk}, \sigma_{E_k,\mathrm{w}}^2 \sim N(\mu_{gk}, \sigma_{E_k,\mathrm{w}}^2)$ in order to guarantee that the k-th study variance of E is $\sigma_{E_k}^2$. Then the first and second moments of E are the same with or without assumption 4.1. Numerically, E has identical marginal distributions with or without assumption 4.1, as it has symmetric distributions under this setting.

B.4 Technical details for chapter 5

Modeling $P(G|E, \boldsymbol{S})$ under HWE

Let $q(E_{ki}, \boldsymbol{S}_{ki})$ be the minor allele frequency for given $(E_{ki}, \boldsymbol{S}_{ki})$. Under the HWE conditional on $(E, \boldsymbol{S})$, we have

$$\begin{aligned} P(G_{ki} = 0|E_{ki}, \boldsymbol{S}_{ki}) &= (1 - q(E_{ki}, \boldsymbol{S}_{ki}))^2, \\ P(G_{ki} = 1|E_{ki}, \boldsymbol{S}_{ki}) &= 2q(E_{ki}, \boldsymbol{S}_{ki})(1 - q(E_{ki}, \boldsymbol{S}_{ki})), \\ P(G_{ki} = 2|E_{ki}, \boldsymbol{S}_{ki}) &= q(E_{ki}, \boldsymbol{S}_{ki})^2. \end{aligned}$$

Then

$$\begin{aligned} \log\left\{\frac{P(G_{ki} = 1|E_{ki}, \boldsymbol{S}_{ki})}{P(G_{ki} = 0|E_{ki}, \boldsymbol{S}_{ki})}\right\} &= \log(2) + \log\left\{\frac{q(E_{ki}, \boldsymbol{S}_{ki})}{1 - q(E_{ki}, \boldsymbol{S}_{ki})}\right\}, \\ \log\left\{\frac{P(G_{ki} = 2|E_{ki}, \boldsymbol{S}_{ki})}{P(G_{ki} = 0|E_{ki}, \boldsymbol{S}_{ki})}\right\} &= 2\log\left\{\frac{q(E_{ki}, \boldsymbol{S}_{ki})}{1 - q(E_{ki}, \boldsymbol{S}_{ki})}\right\}. \end{aligned}$$

We can model the MAF as

$$q(E_{ki}, \boldsymbol{S}_{ki}) = H\{\eta_{0k} + \eta_k \boldsymbol{S}_{ki}^\top + \theta_k E_{ki}\},$$

which can be reduced to

$$q(E_{ki}, \boldsymbol{S}_{ki}) = H\{\eta_{0k} + \eta_k \boldsymbol{S}_{ki}^\top\}$$

under G-E independence conditional on $\boldsymbol{S}_{ki}$.

Chain rule of derivatives: $\hat{\Delta} = \partial\hat{\boldsymbol{\beta}}^\top(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\mathbf{0}} \approx \boldsymbol{I}_{\theta\beta}\boldsymbol{I}_{\beta\beta}^{-1}|_{\boldsymbol{\theta}=\mathbf{0}}$

Using chain rule of derivatives, we have

$$\frac{\partial\ell}{\partial\boldsymbol{\theta}} = \frac{\partial\ell}{\partial\boldsymbol{\beta}}\frac{\partial\boldsymbol{\beta}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}, \quad \text{and then} \quad \left\{\frac{\partial\boldsymbol{\beta}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\}^\top = \left\{\frac{\partial\ell}{\partial\boldsymbol{\theta}^\top}\frac{\partial\ell}{\partial\boldsymbol{\beta}}\right\}\left\{\frac{\partial\ell}{\partial\boldsymbol{\beta}^\top}\frac{\partial\ell}{\partial\boldsymbol{\beta}}\right\}^{-1}.$$

Let $\boldsymbol{I}_{\theta\beta} = -\frac{\partial^2\ell}{\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\beta}}$ and $\boldsymbol{I}_{\beta\beta} = -\frac{\partial^2\ell}{\partial\boldsymbol{\beta}^\top\partial\boldsymbol{\beta}}$ be the corresponding sub-matrices of the full observed information matrix under the unconstrained model. Since $E(\frac{\partial\ell}{\partial\boldsymbol{\theta}^\top}\frac{\partial\ell}{\partial\boldsymbol{\beta}}) = -E(\frac{\partial^2\ell}{\partial\boldsymbol{\theta}^\top\partial\boldsymbol{\beta}})$

and $E(\frac{\partial \ell}{\partial \boldsymbol{\beta}^\top}\frac{\partial \ell}{\partial \boldsymbol{\beta}}) = -E(\frac{\partial^2 \ell}{\partial \boldsymbol{\beta}^\top \partial \boldsymbol{\beta}})$, we use $\boldsymbol{I}_{\theta\beta}\boldsymbol{I}_{\beta\beta}^{-1}|_{\boldsymbol{\theta}=\mathbf{0}}$ as an approximation to the estimator $\hat{\Delta} = \partial \hat{\boldsymbol{\beta}}^\top(\boldsymbol{\theta})/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\mathbf{0}}$.

EB estimator $\hat{\boldsymbol{A}}$

We consider the EB estimator $\hat{\boldsymbol{A}}$ using the following strategies: (1) a conservative estimator $\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top$ for $\boldsymbol{A}$; (2) $\hat{\boldsymbol{A}} = diag(\hat{\tau}^2, ..., \hat{\tau}^2)_{(K\times K)}$, where $\hat{\tau}^2$ is the MLE that maximizes the marginal likelihood of $\hat{\boldsymbol{\theta}}$; (3) $\hat{\boldsymbol{A}} = diag(\bar{\tau}^2, ..., \bar{\tau}^2)_{(K\times K)}$, where $\bar{\tau}^2$ is the estimated posterior mean of $\tau^2|\hat{\boldsymbol{\theta}}$.

(1) and (2) Note that $\hat{\theta}_k \sim N(0, \tau^2 + \hat{\sigma}^2_{\theta_k})$ after marginalizing over θ_k. clearly, $\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^\top$ can serve as a conservative estimator for $\boldsymbol{A}$. One could alternatively estimate τ^2 by maximizing the marginal likelihood of $\hat{\theta}_k$,

$$m(\hat{\theta}_k|\,\tau^2) = \prod_{k=1}^{K}\{2\pi(\tau^2 + \hat{\sigma}^2_{\theta_k})\}^{-\frac{1}{2}} \exp\big\{-\frac{\hat{\theta}_k^2}{2(\tau^2 + \hat{\sigma}^2_{\theta_k})}\big\}.$$

We have

$$\frac{d\,\log\{m(\hat{\theta}_k|\,\tau^2)\}}{d\,\tau^2} = -\frac{1}{2}\sum_{k=1}^{K}\big\{\frac{1}{\tau^2 + \hat{\sigma}^2_{\theta_k}} - \frac{\hat{\theta}_k^2}{(\tau^2 + \hat{\sigma}^2_{\theta_k})^2}\big\},$$

then

$$\frac{d\,\log\{m(\hat{\theta}_k|\,\tau^2)\}}{d\,\tau^2} = 0 \quad \Rightarrow \quad \sum_{k=1}^{K}\frac{\tau^2 - (\hat{\theta}_k^2 - \hat{\sigma}^2_{\theta_k})}{(\tau^2 + \hat{\sigma}^2_{\theta_k})^2} = 0$$

$$\Rightarrow \quad \hat{\tau}^2 = \frac{\sum_{k=1}^{K}(\hat{\theta}_k^2 - \hat{\sigma}^2_{\theta_k})(\hat{\tau}^2 + \hat{\sigma}^2_{\theta_k})^{-2}}{\sum_{k=1}^{K}(\hat{\tau}^2 + \hat{\sigma}^2_{\theta_k})^{-2}}$$

$\hat{\tau}^2$ does not have a closed form expression, however, we could implement an iterative scheme to calculate $\hat{\tau}^2$. If the convergence is to a negative value of $\hat{\tau}^2$, then the MLE of τ^2 is probably 0 Berger (1985). If one assumes $\theta_k \overset{iid}{\sim} N(0, \hat{\tau}^2)$ and proceeds with a Bayesian analysis that ignores the fact $\hat{\tau}^2$ being estimated, then the errors introduced in the hyperparameter estimation would not be reflected in the inference.

(3) Let $B_k = \frac{\hat{\sigma}^2_{\theta_k}}{\tau^2+\hat{\sigma}^2_{\theta_k}}$. The posterior distribution $\pi(\theta_k|\hat{\theta}_k, B_k)$ for given B_k (or equivalently given τ^2) is $N(\mu_{k,\,B_k},\, V_{k,\,B_k})$, where

$$\begin{aligned} \mu_{k,\,B_k} &= (1-B_k)\hat{\theta}_k, \\ V_{k,\,B_k} &= \frac{\tau^2\hat{\sigma}^2_{\theta_k}}{\tau^2+\hat{\sigma}^2_{\theta_k}} = (1-B_k)\hat{\sigma}^2_{\theta_k}. \end{aligned}$$

Rubin (1981), Morris (1983a and 1983b), Laird and Louis (1987) all introduced a vague hyperprior distribution on τ^2, say $H(\tau^2)$, to derive the marginal posterior distribution of $\theta_k|\hat{\theta}_k$ as

$$m(\theta_k|\hat{\boldsymbol{\theta}}) = \int \pi(\theta_k|\hat{\boldsymbol{\theta}}, B_k) dH(B_k|\hat{\boldsymbol{\theta}}),$$

where $H(B_k|\hat{\boldsymbol{\theta}})$ is the posterior distribution of the hyperparameters B_k given $\hat{\boldsymbol{\theta}}$. The marginal posterior mean μ_k and variance V_k of $\theta_k|\hat{\boldsymbol{\theta}}$ are

$$\begin{aligned} \mu_k &= E(\mu_{k,\,B_k}) = (1-E(B_k))\hat{\theta}_k, \\ V_k &= E(V_{k,\,B_k}) + var(\mu_{k,\,B_k}) = (1-E(B_k))\hat{\sigma}^2_{\theta_k} + var(B_k)\hat{\theta}^2_k, \end{aligned}$$

where $E(\cdot)$ and $var(\cdot)$ are taken with respect to $H(B_k|\hat{\boldsymbol{\theta}})$.

To avoid integration in $E(B_k)$ and $var(B_k)$, Morris (1983a and 1983b) used the following estimated posterior mean and variance of B_k (for $K \geq 3$, under a flat prior on τ^2)

$$\begin{aligned} \hat{B}_k &= \frac{K-2}{K}\frac{\hat{\sigma}^2_{\theta_k}}{\hat{\tau}^2+\hat{\sigma}^2_{\theta_k}}, \\ \hat{var}(\hat{B}_k) &= \frac{2}{K-2}\hat{B}^2_k\frac{\hat{\tau}^2+\overline{\sigma}^2_{\theta_k}}{\hat{\tau}^2+\hat{\sigma}^2_{\theta_k}} \end{aligned}$$

for $E(B_k)$ and $var(B_k)$ respectively, where $\hat{\tau}^2$ is the MLE derived under $m(\hat{\theta}_k|\,\tau^2)$ and

$$\overline{\sigma}^2_{\theta_k} = \frac{\sum_{k=1}^K \hat{\sigma}^2_{\theta_k}/(\hat{\tau}^2+\hat{\sigma}^2_{\theta_k})}{\sum_{k=1}^K 1/(\hat{\tau}^2+\hat{\sigma}^2_{\theta_k})}.$$

The estimated posterior mean of $\tau^2|\hat{\boldsymbol{\theta}}$, $\bar{\tau}^2$, can be calculated accordingly. The EB estimators for μ_k and V_k are

$$\hat{\mu}_k = (1-\hat{B}_k)\,\hat{\theta}_k \quad \text{and} \quad \hat{V}_k = (1-\hat{B}_k)\,\hat{\sigma}^2_{\theta_k} + \frac{2}{K-2}\,\hat{B}_k^2\,\hat{\theta}_k^2\,\frac{\hat{\tau}^2+\overline{\sigma}^2_{\theta_k}}{\hat{\tau}^2+\hat{\sigma}^2_{\theta_k}}.$$

Note that $\hat{\mu}_k$ reduced to the James-Stein estimator of θ_k if all $\hat{\sigma}^2_{\theta_k}$ are equal (Morris 1983). The factor $\frac{K-2}{K}$ in $\hat{B}_k$ is used to adjust for the error in the estimation of $\hat{\tau}^2$. When K is large, $\hat{B}_k \to \frac{\hat{\sigma}^2_{\theta_k}}{\hat{\tau}^2+\hat{\sigma}^2_{\theta_k}}$ and $v\hat{a}r(\hat{B}_k) \to 0$, thus $\hat{\mu}_k = \frac{\hat{\tau}^2}{\hat{\tau}^2+\hat{\sigma}^2_{\theta_k}}\,\hat{\theta}_k$ and $\hat{V}_k = \frac{\hat{\tau}^2\hat{\sigma}^2_{\theta_k}}{\hat{\tau}^2+\hat{\sigma}^2_{\theta_k}}$. So a naive analysis by plugging in the estimated prior via MLE $\hat{\tau}^2$ in $N(\mu_{k,\,B_k},\, V_{k,\,B_k})$ will work well when K is large. When K is small or moderate, one must take into account the uncertainty in $\hat{\tau}^2$ while obtaining the posterior distribution of θ_k in the EB spirit (Morris (1983), Berger (1985)).

Proof of lemma 5.2

(i) Proof of $\tilde{\theta}_k = \hat{\theta}_k$: Let $\ell_k(\boldsymbol{\beta}_k, \theta_k) = \log(L_k^R(\boldsymbol{\beta}_k, \theta_k))$ be the log retrospective likelihood of the unconstrained model corresponding to the k-study. Let $\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_k \ell_k(\boldsymbol{\beta}_k, \theta_k)$. We have

$$\max_{\boldsymbol{\theta}}\ \ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = \max_{\boldsymbol{\theta}}\ \sum_k \ell_k(\boldsymbol{\beta}_k, \theta_k) = \sum_k \max_{\theta_k}\ \ell_k(\boldsymbol{\beta}_k, \theta_k).$$

Since the MLE for $\boldsymbol{\theta}$ under L^R is unique, we have

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = (\operatorname*{argmax}_{\theta_1}\, \ell_1(\boldsymbol{\beta}_1, \theta_1), ..., \operatorname*{argmax}_{\theta_K}\, \ell_K(\boldsymbol{\beta}_K, \theta_K)) = \tilde{\boldsymbol{\theta}}.$$

(ii) Proof of $\tilde{\sigma}^2_{\theta_k} \geq \hat{\sigma}^2_{\theta_k}$: Let $\boldsymbol{\zeta}_k = (\gamma_{0k}, \eta_{0k}, \eta_k)$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, ..., \boldsymbol{\zeta}_K)$. It follows that $\boldsymbol{\beta}_k = (\boldsymbol{\gamma}, \boldsymbol{\zeta}_k)$ and $\boldsymbol{\beta} = (\boldsymbol{\gamma}, \boldsymbol{\zeta})$. Note that the derivative of the log likelihood ℓ, with respect to any of the nuisance parameter $\boldsymbol{\zeta}_k$ or θ_k, only depends on the k-study likelihood ℓ_k, i.e.,

$$\frac{\partial \ell}{\partial \theta_k} = \frac{\partial \sum_k \ell_k}{\partial \theta_k} = \frac{\partial \ell_k}{\partial \theta_k} \quad \text{and} \quad \frac{\partial \ell}{\partial \boldsymbol{\zeta}_k} = \frac{\partial \ell_k}{\partial \boldsymbol{\zeta}_k}.$$

Then we have

$$\frac{\partial^2 \ell}{\partial \theta_k^2} = \frac{\partial^2 \ell_k}{\partial \theta_k^2} \text{ and } \frac{\partial^2 \ell}{\partial \theta_j \theta_k} = 0 \ (j \neq k) \quad \Longrightarrow \quad \boldsymbol{I}_{\theta\theta} = diag(\boldsymbol{I}_{1\theta_1\theta_1}, ..., \boldsymbol{I}_{K\theta_K\theta_K})$$

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\gamma}^\top \boldsymbol{\gamma}} = \sum_k \frac{\partial^2 \ell_k}{\partial \boldsymbol{\gamma}^\top \boldsymbol{\gamma}} \quad \Longrightarrow \quad \boldsymbol{I}_{\gamma\gamma} = \sum_k \boldsymbol{I}_{k\gamma\gamma}$$

$$\frac{\partial^2 \ell}{\partial \boldsymbol{\gamma}^\top \theta_k} \frac{\partial^2 \ell}{\partial \theta_k \boldsymbol{\gamma}} = \frac{\partial^2 \ell_k}{\partial \boldsymbol{\gamma}^\top \theta_k} \frac{\partial^2 \ell_k}{\partial \theta_k \boldsymbol{\gamma}} \quad \Longrightarrow \quad \boldsymbol{I}_{\gamma\theta} \boldsymbol{I}_{\theta\gamma} = \sum_k \boldsymbol{I}_{k\gamma\theta_k} \boldsymbol{I}_{k\theta_k\gamma}$$

Similarly $\boldsymbol{I}_{\zeta\zeta} = diag(\boldsymbol{I}_{1\zeta_1\zeta_1}, ..., \boldsymbol{I}_{K\zeta_K\zeta_K})$ and $\boldsymbol{I}_{\theta_j\zeta_k} = 0$ for $j \neq k$. Note that

$$\boldsymbol{I}_{\beta\beta} = \begin{pmatrix} \boldsymbol{I}_{\gamma\gamma} & \boldsymbol{I}_{\gamma\zeta} \\ \boldsymbol{I}_{\zeta\gamma} & \boldsymbol{I}_{\zeta\zeta} \end{pmatrix} = \begin{pmatrix} \sum_k \boldsymbol{I}_{k\gamma\gamma} & (\boldsymbol{I}_{\gamma\zeta_1}, ..., \boldsymbol{I}_{\gamma\zeta_K}) \\ (\boldsymbol{I}_{\gamma\zeta_1}, ..., \boldsymbol{I}_{\gamma\zeta_K})^\top & diag(\boldsymbol{I}_{1\zeta_1\zeta_1}, ..., \boldsymbol{I}_{K\zeta_K\zeta_K}) \end{pmatrix}.$$

Then $\boldsymbol{I}_{\beta\beta}$ can be considered as the summation of a sequence of matrices $\boldsymbol{C}_k$ defined as below, i.e., $\boldsymbol{I}_{\beta\beta} = \sum_k \boldsymbol{C}_k$.

$$\boldsymbol{C}_k = \begin{pmatrix} \boldsymbol{I}_{k\gamma\gamma} & \mathbf{0} & \cdots & \boldsymbol{I}_{k\gamma\zeta_k} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & & & & \\ \vdots & & \ddots & & & \\ \boldsymbol{I}_{k\zeta_k\gamma} & & & \boldsymbol{I}_{k\zeta_k\zeta_k} & & \\ \vdots & & & & \ddots & \\ \mathbf{0} & & & & & \mathbf{0} \end{pmatrix} \qquad \boldsymbol{I}_{k\beta_k\beta_k} = \begin{pmatrix} \boldsymbol{I}_{k\gamma\gamma} & \boldsymbol{I}_{k\gamma\zeta_k} \\ \boldsymbol{I}_{k\zeta_k\gamma} & \boldsymbol{I}_{k\zeta_k\zeta_k} \end{pmatrix}.$$

In above, all the unmarked elements in $\boldsymbol{C}_k$ are zero. $\boldsymbol{C}_k$ only has four nonzero blocks (corresponding to position $(1,1)$, $(1,k+1)$, $(k+1,1)$ and $(k+1,k+1)$ in terms of block), which are the same as the blocks in $\boldsymbol{I}_{k\beta_k\beta_k}$. Clearly, $\boldsymbol{C}_k$'s only have nonnegative eigenvalues: positive eigenvalues identical to that of $\boldsymbol{I}_{k\beta_k\beta_k}$ and zero eigenvalues elsewhere. Thus, all $\boldsymbol{C}_k$'s are symmetric positive semi-definite.

Since $(\boldsymbol{I}^{\theta\theta})^{-1} = \boldsymbol{I}_{\theta\theta} - \boldsymbol{I}_{\theta\beta}\boldsymbol{I}_{\beta\beta}^{-1}\boldsymbol{I}_{\beta\theta}$, we can write

$$\begin{aligned}
\hat{\sigma}_{\theta_k}^{-2} &= \boldsymbol{I}_{k\theta_k\theta_k} - \boldsymbol{I}_{\theta_k\beta}\boldsymbol{I}_{\beta\beta}^{-1}\boldsymbol{I}_{\beta\theta_k} \\
&= \boldsymbol{I}_{k\theta_k\theta_k} - (\boldsymbol{I}_{\theta_k\gamma}, \boldsymbol{I}_{\theta_k\zeta_1}, ..., \boldsymbol{I}_{\theta_k\zeta_K})\boldsymbol{I}_{\beta\beta}^{-1}(\boldsymbol{I}_{\theta_k\gamma}, \boldsymbol{I}_{\theta_k\zeta_1}, ..., \boldsymbol{I}_{\theta_k\zeta_K})^\top \\
&= \boldsymbol{I}_{k\theta_k\theta_k} - (\boldsymbol{I}_{k\theta_k\gamma}, \boldsymbol{0}, ..., \boldsymbol{I}_{k\theta_k\zeta_k}, ..., \boldsymbol{0})(\sum_k \boldsymbol{C}_k)^{-1}(\boldsymbol{I}_{k\theta_k\gamma}, \boldsymbol{0}, ..., \boldsymbol{I}_{k\theta_k\zeta_k}, ..., \boldsymbol{0})^\top \\
\text{and} \quad \tilde{\sigma}_{\theta_k}^{-2} &= \boldsymbol{I}_{k\theta_k\theta_k} - \boldsymbol{I}_{k\theta_k\beta_k}\boldsymbol{I}_{k\beta_k\beta_k}^{-1}\boldsymbol{I}_{k\beta_k\theta_k} \\
&= \boldsymbol{I}_{k\theta_k\theta_k} - (\boldsymbol{I}_{k\theta_k\gamma}, \boldsymbol{I}_{k\theta_k\zeta_k})\boldsymbol{I}_{k\beta_k\beta_k}^{-1}(\boldsymbol{I}_{k\theta_k\gamma}, \boldsymbol{I}_{k\theta_k\zeta_k})^\top. \\
&= \boldsymbol{I}_{k\theta_k\theta_k} - (\boldsymbol{I}_{k\theta_k\gamma}, \boldsymbol{0}, ..., \boldsymbol{I}_{k\theta_k\zeta_k}, ..., \boldsymbol{0})\boldsymbol{C}_k^{-1}(\boldsymbol{I}_{k\theta_k\gamma}, \boldsymbol{0}, ..., \boldsymbol{I}_{k\theta_k\zeta_k}, ..., \boldsymbol{0})^\top.
\end{aligned}$$

In order to show $\tilde{\sigma}_{\theta_k}^2 \geq \hat{\sigma}_{\theta_k}^2$ for $k = 1, ..., K$, it is sufficient to show $\boldsymbol{C}_k^{-1} \geq (\sum_k \boldsymbol{C}_k)^{-1}$ for each k. It is equivalent to show $\sum_{j\neq k} \boldsymbol{C}_j \geq 0$ for each k, which is true because all $\boldsymbol{C}_k$'s are symmetric positive semi-definite.

The equality in $\tilde{\sigma}_{\theta_k}^2 \geq \hat{\sigma}_{\theta_k}^2$ hold iff $\boldsymbol{C}_j = 0$ for all $j \neq k$. The symmetric positive semi-definite matrix $\boldsymbol{C}_j = 0$ iff $\boldsymbol{I}_{k\beta_k\beta_k} = 0$, i.e., ℓ_k contains no information. Therefore, the inequality in $\tilde{\sigma}_{\theta_k}^2 \geq \hat{\sigma}_{\theta_k}^2$ is usually strict in practice. Although each $\tilde{\theta}_k$ is derived under individual likelihood L_k^R and $\hat{\theta}_k$'s are derived jointly under L^R, we actually have $\tilde{\theta}_k = \hat{\theta}_k$ and $\tilde{\sigma}_{\theta_k}^2 \geq \hat{\sigma}_{\theta_k}^2$.

Derivation of $\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}_{EB}}$

(1) $cov(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0)$: Denote $\boldsymbol{I}$ ($\boldsymbol{I}^0$) as the full observed information matrix for the unconstrained (constrained) model; denote $U_{ki}(\boldsymbol{\beta}, \boldsymbol{\theta})$ and $U_{ki}^0(\boldsymbol{\beta})$ as the individual score functions for subject (k, i) for the unconstrained and constrained models respectively. Asymptotically,

$$\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \end{pmatrix} = n^{\frac{1}{2}}\boldsymbol{I}^{-1}\sum_{k,i} U_{ki}(\boldsymbol{\beta}, \boldsymbol{\theta}) \text{ and } \sqrt{n}(\hat{\boldsymbol{\beta}}(\boldsymbol{0}) - \boldsymbol{\beta}(\boldsymbol{0})) = n^{\frac{1}{2}}(\boldsymbol{I}^0)^{-1}\sum_{k,i} U_{ki}^0(\boldsymbol{\beta}),$$

where $\boldsymbol{\beta}^*$ and $\boldsymbol{\theta}^*$ are the true parameter values, and $\boldsymbol{\beta}(\mathbf{0})$ denotes the limiting value of the profile MLE $\hat{\boldsymbol{\beta}}(\mathbf{0})$. The joint asymptotic variance-covariance matrix $\boldsymbol{\Sigma}_{\hat{\Omega}}$ of $\hat{\boldsymbol{\Omega}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}^0)$ is

$$\begin{pmatrix} \boldsymbol{I}^{-1} var(\sum_{k,i} U_{ki}) \boldsymbol{I}^{-1} & \boldsymbol{I}^{-1} cov(\sum_{k,i} U_{ki}, \sum_{k,i} U^0_{ki}) (\boldsymbol{I}^0)^{-1} \\ (\boldsymbol{I}^0)^{-1} cov(\sum_{k,i} U^0_{ki}, \sum_{k,i} U_{ki}) \boldsymbol{I}^{-1} & (\boldsymbol{I}^0)^{-1} var(\sum_{k,i} U^0_{ki}) (\boldsymbol{I}^0)^{-1} \end{pmatrix}.$$

Then $cov(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0)$ can be obtained as the corresponding sub-matrix of $\boldsymbol{\Sigma}_{\hat{\Omega}}$.

(2) $\hat{\boldsymbol{V}}_{\hat{\gamma}_{EB1}}$: We consider $\hat{\boldsymbol{\gamma}}_{\text{EB1}} = \hat{\boldsymbol{\psi}}\hat{\boldsymbol{\psi}}^\top \{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}} + \hat{\boldsymbol{\psi}}\hat{\boldsymbol{\psi}}^\top\}^{-1} \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}} \{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}} + \hat{\boldsymbol{\psi}}\hat{\boldsymbol{\psi}}^\top\}^{-1} \hat{\boldsymbol{\gamma}}^0 = \boldsymbol{f}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0)$ as a function of $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0)$. Applying Delta method, the approximate variance-covariance matrix of $\hat{\boldsymbol{\gamma}}_{\text{EB1}}$ is given by $\boldsymbol{f}'(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0)^\top cov(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0) \boldsymbol{f}'(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0)$, where $\boldsymbol{f}'$ is the $p \times 2p$ gradient matrix that can be derived as

$$\boldsymbol{f}' = \Big(-\frac{2\hat{\boldsymbol{\psi}}\hat{\boldsymbol{\psi}}^\top \hat{\boldsymbol{V}}^{-1}_{\hat{\boldsymbol{\gamma}}}}{(1 + \hat{\boldsymbol{\psi}}^\top \hat{\boldsymbol{V}}^{-1}_{\hat{\boldsymbol{\gamma}}} \hat{\boldsymbol{\psi}})^2} + \frac{\mathbf{1}_p}{1 + \hat{\boldsymbol{\psi}}^\top \hat{\boldsymbol{V}}^{-1}_{\hat{\boldsymbol{\gamma}}} \hat{\boldsymbol{\psi}}}, \quad \frac{2\hat{\boldsymbol{\psi}}\hat{\boldsymbol{\psi}}^\top \hat{\boldsymbol{V}}^{-1}_{\hat{\boldsymbol{\gamma}}}}{(1 + \hat{\boldsymbol{\psi}}^\top \hat{\boldsymbol{V}}^{-1}_{\hat{\boldsymbol{\gamma}}} \hat{\boldsymbol{\psi}})^2} + \frac{\hat{\boldsymbol{\psi}}^\top \hat{\boldsymbol{V}}^{-1}_{\hat{\boldsymbol{\gamma}}} \hat{\boldsymbol{\psi}} \mathbf{1}_p}{1 + \hat{\boldsymbol{\psi}}^\top \hat{\boldsymbol{V}}^{-1}_{\hat{\boldsymbol{\gamma}}} \hat{\boldsymbol{\psi}}} \Big)$$

(3) $\hat{\boldsymbol{V}}_{\hat{\gamma}_{EB2}}$ and $\hat{\boldsymbol{V}}_{\hat{\gamma}_{EB3}}$: For $\hat{\boldsymbol{\gamma}}_{\text{EB2}}$ and $\hat{\boldsymbol{\gamma}}_{\text{EB3}}$ of the more general form $\hat{\boldsymbol{\gamma}}_{\text{EB}} = \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta} \{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}} + \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta}\}^{-1} \hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}} \{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}} + \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta}\}^{-1} \hat{\boldsymbol{\gamma}}^0$, it is hard to apply the above Delta method to obtain a variance formula for the EB estimator since $\hat{\boldsymbol{A}}$ does not have a closed form expression in terms of $\hat{\boldsymbol{\Omega}}$. Instead, we consider an *ad hoc* way to calculate the variance. Let $\boldsymbol{w} = \Delta^\top \boldsymbol{A} \Delta \{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}} + \Delta^\top \boldsymbol{A} \Delta\}^{-1}$. We derive the functional form of $\boldsymbol{V}_{\hat{\gamma}_{EB}}$ treating $\boldsymbol{w}$ as fixed, i.e., $\hat{\boldsymbol{\gamma}}_{\text{EB}} = \boldsymbol{w}\hat{\boldsymbol{\gamma}} + (\mathbf{1}_p - \boldsymbol{w})\hat{\boldsymbol{\gamma}}^0$, and then plug in the estimated weight $\hat{\boldsymbol{w}} = \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta} \{\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}} + \hat{\Delta}^\top \hat{\boldsymbol{A}} \hat{\Delta}\}^{-1}$ into the functional form to get $\hat{\boldsymbol{V}}_{\hat{\gamma}_{EB}}$. In particular, $\boldsymbol{V}_{\hat{\gamma}_{EB}} = \boldsymbol{w}\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}}\boldsymbol{w}^\top + (\mathbf{1}_p - \boldsymbol{w})\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}^0}(\mathbf{1}_p - \boldsymbol{w})^\top + 2\boldsymbol{w} cov(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0)(\mathbf{1}_p - \boldsymbol{w})^\top$. Then the *ad hoc* estimator $\hat{\boldsymbol{V}}_{\hat{\gamma}_{EB}} = \hat{\boldsymbol{w}}\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}}\hat{\boldsymbol{w}}^\top + (\mathbf{1}_p - \hat{\boldsymbol{w}})\hat{\boldsymbol{V}}_{\hat{\boldsymbol{\gamma}}^0}(\mathbf{1}_p - \hat{\boldsymbol{w}})^\top + 2\hat{\boldsymbol{w}} \hat{cov}(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\gamma}}^0)(\mathbf{1}_p - \hat{\boldsymbol{w}})^\top$.

A mixture distribution: $\theta_k \sim p\,\delta(0) + (1-p)N(0, \tau^2)$

For the mixture distribution $\theta_k \sim p\,\delta(0) + (1-p)N(0, \tau^2)$, we have $E(\theta_k) = 0$ and var$(\theta_k) = (1-p)\tau^2$ (This can be proved via $E(\theta_k) = E[E(\theta_k|L)]$ and $Var(X) =$

$E[Var(\theta_k|L)] + Var[E(\theta_k|L)]$). For a latent random variable $L \sim Bin(p)$, $\theta_k|L = 1 \sim \delta(0)$; $\theta_k|L = 0 \sim N(0, \tau^2)$. Marginalizing over θ_k, it follows that $\hat{\theta}_k|L = 1 \sim N(0, \hat{\sigma}^2_{\theta_k})$; $\hat{\theta}_k|L = 0 \sim N(0, \hat{\sigma}^2_{\theta_k} + \tau^2)$. So marginally, $\hat{\theta}_k$ has a mixture distribution $\hat{\theta}_k \sim pN(0, \hat{\sigma}^2_{\theta_k}) + (1-p)N(0, \hat{\sigma}^2_{\theta_k} + \tau^2)$, which has mean 0 and variance $\hat{\sigma}^2_{\theta_k} + (1-p)\tau^2$. Similarly, we estimate the prior hyperparameter p and τ^2 by maximizing the marginal likelihood of $\hat{\theta}_k$

$$m(\hat{\theta}_k|p, \tau^2) = \prod_{k=1}^{K} \Big[p(2\pi\hat{\sigma}^2_{\theta_k})^{-\frac{1}{2}} \exp\big\{ -\frac{\hat{\theta}^2_k}{2\hat{\sigma}^2_{\theta_k}} \big\} + (1-p)\{2\pi(\tau^2 + \hat{\sigma}^2_{\theta_k})\}^{-\frac{1}{2}} \exp\big\{ -\frac{\hat{\theta}^2_k}{2(\tau^2 + \hat{\sigma}^2_{\theta_k})} \big\} \Big].$$

Denote the MLE of p and τ^2 by $\hat{p}$ and $\hat{\tau}^2$.

BIBLIOGRAPHY

BIBLIOGRAPHY

Agresti, A. (2002). *Categorical data analysis*. Wiley-interscience.

Antoniak, C. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics* **2,** 1152–1174.

Aschard, H., Hancock, D. B., London, S. J., and Kraft, P. (2011). Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Human Heredity* **70,** 292–300.

Baddeley, A., Gelfand, A., Diggle, P., Fuentes, M., and Guttorp, P. (2010). Modelling strategies. *Handbook of Spatial Statistics* pages 339–369.

Baker, S. (1994). The multinomial-Poisson transformation. *The Statistician* **1,** 495–504.

Banerjee, S., Gelfand, A. E., and Carlin, B. P. (2003). *Hierarchical modeling and analysis for spatial data*. Crc Press.

Basu, R., Dominici, F., and Samet, J. (2005). Temperature and mortality among the elderly in the United States: A comparison of epidemiologic methods. *Epidemiology* **16,** 58–66.

Bateson, T. F. and Schwartz, J. (1999). Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. *Epidemiology* **10,** 539–544.

Batterman, S. A., Zhang, K., Kononowech, R., et al. (2010). Prediction and analysis of near-road concentrations using a reduced-form emission/dispersion model. *Environmental Health* **9,** 29–29.

Becher, H. (1991). Alternative parameterization of polychotomous models: Theory and application to matched case-ontrol studies. *Statistics in Medicine* **10,** 375–382.

Becher, K. and Jöckel, K. (1990). Bias adjustment with polychotomous logistic regression in matched case-control studies with two control groups. *Biometrical Journal* **32,** 801–816.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer Verlag.

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

Bhattacharjee, S., Chatterjee, N., Han, S., and Wheeler, W. (2012). *CGEN: An R package for analysis of case-control studies in genetic epidemiology*. R package version 2.2.0.

Blackwell, D. and MacQueen, J. (1973). Ferguson distributions via Pólya urn schemes. *The Annals of Statistics* **1,** 353–355.

Breslow, N. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **97,** 9–25.

Breslow, N., Day, N., et al. (1980). *Statistical methods in cancer research. Vol. 1. The analysis of case-control studies.* Distributed for IARC by WHO, Geneva, Switzerland.

Breslow, N., Day, N., Halvorsen, K., Prentice, R., and Sabai, C. (1978). Estimation of multiple relative risk functions in matched case-control studies. *American Journal of Epidemiology* **108,** 299–307.

Carracedo-Martínez, E., Taracido, M., Tobias, A., Saez, M., and Figueiras, A. (2010). Case-crossover analysis of air pollution health effects: A systematic review of methodology and application. *Environmental Health Perspectives* **118,** 1173–1182.

Chatterjee, N. and Carroll, R. J. (2005). Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* **92,** 399–418.

Chen, Y.-H., Chatterjee, N., and Carroll, R. J. (2009). Shrinkage estimators for robust and efficient inference in haplotype-based case-control studies. *Journal of the American Statistical Association* **104,** 220–233.

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

Congdon, P. and Congdon, P. (2003). *Applied Bayesian modelling*. Wiley New York.

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

Dey, D. K. and Liu, J. (2007). A quantitative study of quantile based direct prior elicitation from expert opinion. *Bayesian Analysis* **2,** 137–166.

DiCiccio, T., Kass, R., Raftery, A., and Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association* **92,** 903–915.

Diggle, P. (1990). A point process modeling approach to raised incidence of a rare phenomenon in the vicinity of a pre-specified point. *Journal of the Royal Statistical Society A* **153,** 349–362.

Diggle, P., Elliott, P., Morris, S., and Shaddick, G. (1997). Regression modeling of disease risk in relation to point sources. *Journal of the Royal Statistical Society A* **160,** 491–505.

Diggle, P., Morris, S., and Wakefield, J. (2000). Point-source modeling using matched case-control data. *Biostatistics* **1,** 89–105.

Diggle, P. and Rowlingson, B. (1994). A conditional approach to point process modeling of raised incidence. *Journal of the Royal Statistical Society A* **157,** 433–440.

Dominici, F., Daniels, M., Zeger, S., and Samet, J. (2002). Air pollution and mortality: Estimating regional and national dose-response relationships. *Journal of the American Statistical Association* **97,** 100–111.

Dominici, F., McDermott, A., Zeger, S., and Samet, J. (2003). On the use of generalized additive models in time-series studies of air pollution and health. *American Journal of Epidemiology* **156,** 193–203.

Dominici, F., Samet, J., and Zeger, S. (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: A hierarchical modelling strategy. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **163,** 263–302.

Dreassi, E., Lagazio, C., Maule, M., Magnani, C., Biggeri, A., et al. (2008). Sensitivity analysis of the relationship between disease occurrence and distance from a putative source of pollution. *Geospatial Health* **2,** 263–271.

Dubin, N. and Pasternack, B. (1986). Risk assessment for case-control subgroups by polychotomous logistic regression. *American Journal of Epidemiology* **123,** 1101–1117.

Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A. U., Wheeler, E., Glazer, N. L., Bouatia-Naji, N., Gloyn, A. L., et al. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics* **42,** 105–116.

Ferguson, T. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* **1,** 209–230.

Fleiss, J. (1993). Review papers: The statistical basis of meta-analysis. *Statistical Methods in Medical Research* **2,** 121–145.

Freathy, R., Timpson, N., Lawlor, D., Pouta, A., Ben-Shlomo, Y., Ruokonen, A., Ebrahim, S., Shields, B., Zeggini, E., Weedon, M., et al. (2008). Common variation in the FTO gene alters diabetes-related metabolic traits to the extent expected given its effect on BMI. *Diabetes* **57,** 1419–1426.

Gallant, A. (2009). *Nonlinear statistical models*. Wiley.

Gasparrini, A., Armstrong, B., and Kenward, M. (2010). Distributed lag non-linear models. *Statistics in Medicine* **29,** 2224–2234.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* **7,** 457–472.

Ghosh, M. and Chen, M. (2002). Bayesian inference for matched case-control studies. *Sankhyā: The Indian Journal of Statistics, Series B* **64,** 107–127.

Ghosh, M., Song, J., Forster, J., Mitra, R., and Mukherjee, B. (2012). On the equivalence of posterior inference based on retrospective and prospective likelihoods: Application to a case-control study of colorectal cancer. *Statistics in Medicine* **31,** 2196–2208.

Ghosh, M., Zhang, L., and Mukherjee, B. (2006). Equivalence of posteriors in the Bayesian analysis of the multinomial-Poisson transformation. *Metron-International Journal of Statistics* **64,** 19–28.

Greenland, S. (1987). Quantitative methods in the review of epidemiologic literature. *Epidemiologic Reviews* **9,** 1–30.

HEI (2010). Health effects institute: Traffic-related air pollution: A critical review of the literature on emissions, exposure, and health effects. *Special Report 17* .

Ibrahim, J. G. and Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science* **15,** 46–60.

Janes, H., Sheppard, L., and Lumley, T. (2005a). Case-crossover analyses of air pollution exposure data: Referent selection strategies and their implications for bias. *Epidemiology* **16,** 717–726.

Janes, H., Sheppard, L., and Lumley, T. (2005b). Overlap bias in the case-crossover design, with application to air pollution exposures. *Statistics in Medicine* **24,** 285–300.

Kass, R. and Raftery, A. (1995). Bayes factors. *Journal of the american statistical association* **90,** 773–795.

Khoury, M. and Wacholder, S. (2009). Invited commentary: From genome-wide association studies to gene-environment-wide interaction studies: Challenges and opportunities. *American Journal of Epidemiology* **169,** 227–230.

Kilpeläinen, T., Qi, L., Brage, S., Sharp, S., Sonestedt, E., Demerath, E., Ahmad, T., Mora, S., Kaakinen, M., Sandholt, C., et al. (2011). Physical activity attenuates the influence of FTO variants on obesity risk: A meta-analysis of 218,166 adults and 19,268 children. *PLoS Medicine* **8,** e1001116.

Kovalchik, S. A. (2013). Aggregate-data estimation of an individual patient data linear random effects meta-analysis with a patient covariate-treatment interaction term. *Biostatistics* **14,** 273–283.

Laird, N. M. and Louis, T. A. (1987). Empirical Bayes confidence intervals based on bootstrap samples. *Journal of the American Statistical Association* **82,** 739–750.

Lawson, A. (1993). On the analysis of mortality events associated with a pre-specified fixed point. *Journal of the Royal Statistical Society A* **156,** 363–377.

Lawson, A., Browne, W., and Rodeiro, C. (2003). *Disease mapping with WinBUGS and MLwiN*. Wiley.

Levy, D., Lumley, T., Sheppard, L., Kaufman, J., and Checkoway, H. (2001). Referent selection in case-crossover analyses of acute health effects of air pollution. *Epidemiology* **12,** 186–192.

Li, D. and Conti, D. V. (2009). Detecting gene-environment interactions using a combined case-only and case-control approach. *American Journal of Epidemiology* **169,** 497–504.

Li, S., Batterman, S., Wasilevich, E., Elasaad, H., Wahl, R., Mukherjee, B., et al. (2011). Asthma exacerbation and proximity of residence to major roads: A population-based matched case-control study among the pediatric Medicaid population in Detroit, Michigan. *Environmental Health* **10,** 34.

Li, S., Batterman, S., Wasilevich, E., Wahl, R., Wirth, J., Su, F., and Mukherjee, B. (2011). Association of daily asthma emergency department visits and hospital admissions with ambient air pollutants among the pediatric Medicaid population in Detroit: Time-series and time-stratified case-crossover analyses with threshold effects. *Environmental Research* **111,** 1137–1147.

Liang, K. and Stewart, W. (1987). Polychotomous logistic regression methods for matched case-control studies with multiple case or control groups. *American Journal of Epidemiology* **125,** 720–730.

Lin, D. and Zeng, D. (2010). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* **97,** 321–332.

Lu, Y. and Zeger, S. (2007). On the equivalence of case-crossover and time-series methods in environmental epidemiology. *Biostatistics* **8,** 337–344.

Lumley, T. and Levy, D. (2000). Bias in the case-crossover design: Implications for studies of air pollution. *Environmetrics* **11,** 689–704.

Luo, X. and Sorock, G. (2008). Analysis of recurrent event data under the case-crossover design with applications to elderly falls. *Statistics in Medicine* **27,** 2890–2901.

Maclure, M. (1991). The case-crossover design: A method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* **133,** 144–153.

Manning, A., LaValley, M., Liu, C., Rice, K., An, P., Liu, Y., Miljkovic, I., Rasmussen-Torvik, L., Harris, T., Province, M., et al. (2011). Meta-analysis of gene-environment interaction: Joint estimation of SNP and SNP×environment regression coefficients. *Genetic Epidemiology* **35,** 11–18.

Meng, X. (2010). Automated bias-variance trade-off: Intuitive inadmissibility or inadmissible intuition? In *Frontiers of Statistical Decision Making and Bayesian Analysis*, pages 95–112. Springer.

Mittleman, M. (2005). Optimal referent selection strategies in case-crossover studies: A settled issue. *Epidemiology* **16,** 715–716.

Morgenstern, H. (1982). Uses of ecologic analysis in epidemiologic research. *American Journal of Public Health* **72,** 1336–1344.

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segre, A. V., Steinthorsdottir, V., Strawbridge, R. J., Khan, H., Grallert, H., Mahajan, A., et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics* **44,** 981–990.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association* **78,** 47–55.

Mukherjee, B., Ahn, J., Gruber, S., and Chatterjee, N. (2012). Testing gene-environment interaction in large-scale case-control association studies: Possible choices and comparisons. *American Journal of Epidemiology* **175,** 177–190.

Mukherjee, B., Ahn, J., Gruber, S. B., Rennert, G., Moreno, V., and Chatterjee, N. (2008). Tests for gene-environment interaction from case-control data: A novel study of type I error, power and designs. *Genetic Epidemiology* **32,** 615–626.

Mukherjee, B., Ahn, J., Liu, I., and Sánchez, B. (2009). On elimination of nuisance parameters in stratified proportional odds model by amalgamating conditional likelihoods. *Statistics in Medicine* **27,** 4950–4971.

Mukherjee, B. and Chatterjee, N. (2008). Exploiting gene-environment independence for analysis of case-control studies: An empirical Bayes-type shrinkage estimator to trade-off between bias and efficiency. *Biometrics* **64,** 685–694.

Mukherjee, B., Liu, I., and Sinha, S. (2007). Analysis of matched case–control data with multiple ordered disease states: Possible choices and comparisons. *Statistics in Medicine* **26,** 3240–3257.

Müller, P. and Quintana, F. (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19,** 95–110.

Murcray, C. E., Lewinger, J. P., and Gauderman, W. J. (2009). Gene-environment interaction in genome-wide association studies. *American Journal of Epidemiology* **169,** 219–226.

Navidi, W. (1998). Bidirectional case-crossover designs for exposures with time trends. *Biometrics* **54,** 596–605.

Navidi, W. and Weinhandl, E. (2002). Risk set sampling for case-crossover designs. *Epidemiology* **13,** 100–105.

Neal, R. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics* **9,** 249–265.

Piegorsch, W. W., Weinberg, C. R., and Taylor, J. A. (1994). Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statistics in Medicine* **13,** 153–162.

Psaty, B., O'Donnell, C., Gudnason, V., Lunetta, K., Folsom, A., Rotter, J., Uitterlinden, A., Harris, T., Witteman, J., Boerwinkle, E., et al. (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circulation: Cardiovascular Genetics* **2,** 73–80.

Rice, K. (2004). Equivalence between conditional and mixture approaches to the Rasch model and matched case-control studies, with applications. *Journal of the American Statistical Association* **99,** 510–522.

Rice, K. (2008). Equivalence between conditional and random-effects likelihoods for pair-matched case-control studies. *Journal of the American Statistical Association* **103,** 385–396.

Rodrigues, A., Diggle, P., and Assuncao, R. (2010). Semi-parametric approach to point source modeling in epidemiology and criminology. *Journal of the Royal Statistical Society C* **59,** 533–542.

Rubin, D. B. (1981). Estimation in parallel randomized experiments. *Journal of Educational and Behavioral Statistics* **6,** 377–401.

Sarwar, N., Butterworth, A., Freitag, D., Gregson, J., Willeit, P., Gorman, D., Gao, P., Saleheen, D., Rendon, A., Nelson, C., et al. (2012). Interleukin-6 receptor pathways in coronary heart disease: A collaborative meta-analysis of 82 studies. *The Lancet* **379,** 1205–1213.

Saxena, R., Saleheen, D., Been, L. F., Garavito, M. L., Braun, T., Bjonnes, A., Young, R., Ho, W. K., Rasheed, A., Frossard, P., et al. (2013). Genome-wide association study identifies a novel locus contributing to type 2 diabetes susceptibility in Sikhs of Punjabi origin from India. *Diabetes* **62,** 1746–1755.

Schwartz, S. (1994). The fallacy of the ecological fallacy: The potential misuse of a concept and the consequences. *American Journal of Public Health* **84,** 819–824.

Scott, L., Mohlke, K., Bonnycastle, L., Willer, C., Li, Y., Duren, W., Erdos, M., Stringham, H., Chines, P., Jackson, A., et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316,** 1341–1345.

Scott, R. A., Lagou, V., Welch, R. P., Wheeler, E., Montasser, M. E., Luan, J., Mägi, R., Strawbridge, R. J., Rehnberg, E., Gustafsson, S., et al. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics* **44,** 991–1005.

Seaman, S. and Richardson, S. (2004). Equivalence of prospective and retrospective models in the Bayesian analysis of case-control studies. *Biometrika* **91,** 15–25.

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

Sinha, S., Mukherjee, B., and Ghosh, M. (2004). Bayesian semiparametric modeling for matched case–control studies with multiple disease states. *Biometrics* **60,** 41–49.

Song, C., Chen, G. K., Millikan, R. C., Ambrosone, C. B., John, E. M., Bernstein, L., Zheng, W., Hu, J. J., Ziegler, R. G., Nyante, S., et al. (2013). A genome-wide scan for breast cancer risk haplotypes among African American women. *PloS One* **8,** e57298.

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Allen, H. L., Lindgren, C. M., Luan, J., Mägi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* **42,** 937–948.

Staicu, A. (2010). On the equivalence of prospective and retrospective likelihood methods in case-control studies. *Biometrika* **97,** 990–996.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature Genetics* **466,** 707–713.

Thomas, D., Goldberg, M., Dewar, R., and Siemiatycki, J. (1986). Statistical methods for relating several exposure factors to several diseases in case-heterogeneity studies. *Statistics in Medicine* **5,** 49–60.

Umbach, D. M. and Weinberg, C. R. (1997). Designing and analysing case-control studies to exploit independence of genotype and exposure. *Statistics in Medicine* **16,** 1731–1743.

VanderWeele, T., Mukherjee, B., and Chen, J. (2012). Sensitivity analysis for interactions under unmeasured confounding. *Statistics in Medicine* **31,** 2552–2564.

Voight, B. F., Scott, L. J., Steinthorsdottir, V., Morris, A. P., Dina, C., Welch, R. P., Zeggini, E., Huth, C., Aulchenko, Y. S., Thorleifsson, G., et al. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics* **42,** 579–589.

Wakefield, J. and Morris, S. (2001). The Bayesian modeling of disease risk in relation to a point source. *Journal of the American Statistical Association* **96,** 77–91.

Welty, L., Peng, R., Zeger, S., and Dominici, F. (2009). Bayesian distributed lag models: Estimating effects of particulate matter air pollution on daily mortality. *Biometrics* **65,** 282–291.

Whitehead, A. and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine* **10,** 1665–1677.

Willer, C., Li, Y., and Abecasis, G. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191.

Zeggini, E., Scott, L., Saxena, R., Voight, B., Marchini, J., Hu, T., de Bakker, P., Abecasis, G., Almgren, P., Andersen, G., et al. (2008). Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* **40,** 638–645.