The Role of Environmental Heterogeneity in Meta-Analysis of Gene–Environment Interactions With Quantitative Traits

Shi Li,[1] Bhramar Mukherjee,[1]* Jeremy M. G. Taylor,[1] Kenneth M. Rice,[2] Xiaoquan Wen,[1] John D. Rice,[1] Heather M. Stringham,[1] and Michael Boehnke[1]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, United States of America; [2]Department of Biostatistics, University of Washington, Seattle, Washington, United States of America

ABSTRACT: With challenges in data harmonization and environmental heterogeneity across various data sources, meta-analysis of gene–environment interaction studies can often involve subtle statistical issues. In this paper, we study the effect of environmental covariate heterogeneity (within and between cohorts) on two approaches for fixed-effect meta-analysis: the standard inverse-variance weighted meta-analysis and a meta-regression approach. Akin to the results in Simmonds and Higgins (2007), we obtain analytic efficiency results for both methods under certain assumptions. The relative efficiency of the two methods depends on the ratio of within versus between cohort variability of the environmental covariate. We propose to use an adaptively weighted estimator (AWE), between meta-analysis and meta-regression, for the interaction parameter. The AWE retains full efficiency of the joint analysis using individual level data under certain natural assumptions. Lin and Zeng (2010a, b) showed that a multivariate inverse-variance weighted estimator retains full efficiency as joint analysis using individual level data, if the estimates with full covariance matrices for all the common parameters are pooled across all studies. We show consistency of our work with Lin and Zeng (2010a, b). Without sacrificing much efficiency, the AWE uses only univariate summary statistics from each study, and bypasses issues with sharing individual level data or full covariance matrices across studies. We compare the performance of the methods both analytically and numerically. The methods are illustrated through meta-analysis of interaction between Single Nucleotide Polymorphisms in *FTO* gene and body mass index on high-density lipoprotein cholesterol data from a set of eight studies of type 2 diabetes.
Genet Epidemiol 38:416–429, 2014. © 2014 Wiley Periodicals, Inc.

KEY WORDS: adaptively weighted estimator; covariate heterogeneity; gene–environment interaction; individual patient data; meta-analysis; meta-regression; power calculation

Introduction

Genome-wide association studies (GWAS) provide tremendous opportunities for large-scale exploration of associations between genetic variants and complex traits. Searching genetic associations based on GWAS has been successfully identifying marginal effects of variants at multiple susceptibility loci for a wide spectrum of complex traits, e.g. type 2 diabetes (T2D) [Morris et al., 2012; Saxena et al., 2013; Scott et al., 2007; Zeggini et al., 2008], cardiovascular outcomes [Psaty et al., 2009; Sarwar et al., 2012] and cancer [Song et al., 2013]. The agnostic discovery strategy of GWAS can be used to detect gene–environment interactions (GEI) that can further characterize the genetic architecture of complex traits through subgroup or joint effects [Mukherjee et al., 2012; Khoury and Wacholder, 2009]. In general, the definition of "environment" can be broad, including demographic factors (age, gender etc.), behavioral factors (smoking, al-cohol consumption, diet, medication use etc.), and external factors (exposure to air pollution, radio-active substances etc.). Complex traits are influenced by both genetic and environmental factors and possibly their interaction, e.g. physical activity appeared to attenuate the effect of fat mass associated *FTO* gene variants on obesity risk [Kilpeläinen et al., 2011]. With limited number of findings on GEIs so far, it is likely that the GEI effects are small to modest, warranting the need for larger sample sizes and collaboration across different study sites for joint or meta-analysis. Many collaborative networks have been formed to share individual or summary level data from multiple GWAS of related traits, e.g. the DIAGRAM (T2D) [Morris et al., 2012; Voight et al., 2010; Zeggini et al., 2008], MAGIC (glucose and insulin-related traits) [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 (e.g. METAL Willer et al., 2010) to implement GWA meta-analysis (GWAMA). However, there are relatively few papers that explore analytical issues for

meta-analysis of GEI (e.g. Aschard et al., 2011; Manning et al., 2011 to name a couple) compared to meta-analysis of marginal genetic associations.

Several meta-analytic techniques largely developed for their use in randomized clinical trials can be adapted 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 "fixed effect model" in the classical literature [Borenstein et al., 2010; Fleiss, 1993; Lin and Zeng, 2010b; Whitehead and Whitehead, 1991] most often refers to a model with fixed and common effect. But in general, "fixed effects model" (in plural) only requires that there are fixed and unrelated effects in each study, regardless of the homogeneity assumption. Effect homogeneity can be tested by the Cochran's Q-test [Cochran, 1954]. In this paper, we consider the fixed and common effect framework for the interaction parameter as in Lin and Zeng [2010b] to derive our analytical results. We comment on this choice as opposed to a general-fixed effects model where the interaction parameter can be different across studies later in the paper.

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 raw data from all studies as IPD analysis (also called mega-analysis in some papers, e.g. Lin and Zeng [2010a]), and classify the methods that combine summary statistics derived from analysis of different studies as meta-analysis. A natural question to ask is how much efficiency gain, if any, can be achieved by analyzing IPD over meta-analysis. Recently, Lin and Zeng [2010b] considered a multivariate IVW (MIVW) estimator under the common effect model. In constructing the MIVW, if the estimates with full covariance matrix for all the common parameters are pooled across studies, then the MIVW is asymptotically equivalent to the IPD estimator. 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 more likely to be available. Lin and Zeng also quantified the efficiency loss of using an univariate IVW (UIVW) versus a MIVW estimator. The results from Lin and Zeng are derived in a very general setting. In this paper, we specifically focus on the estimation and testing of GEI parameter. Our goal is to construct estimator for GEI parameter using only univariate summary statistics, bypassing issues with sharing individual level data or multivariate covariance matrices across studies, without sacrificing much efficiency or incurring increased bias.

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; Kovalchik 2013) have com-

pared aggregate data analysis (e.g. MR) with IPD analysis to detect treatment-biomarker interactions for randomized clinical trials (analogous to gene–environment interactions in our case). Simmonds and Higgins considered three methods IPD, UIVW, and MR and showed that under certain natural assumptions, analytical power formulae to detect interactions can be expressed in terms of total, within and between study sum of squares (TSS, WSS, and BSS, respectively) corresponding to the environmental covariate. In absence of IPD, they recommended using UIVW rather than MR if the WSS exceeds BSS and vice versa. We borrow from their work to derive similar analytical expressions for testing GEI under certain assumptions.

Instead of making a discrete choice between UIVW versus MR, we propose a novel adaptively weighted estimator (AWE) combining UIVW and MR, and archiving the same asymptotic efficiency as the IPD estimator under certain conditions. The AWE has two major advantages over the MIVW estimator shown in the following main text: (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); (2) AWE has less efficiency loss compared to MIVW under model misspecification, for example, when the main effects of G or E are heterogeneous across studies or when a continuous covariate E is centered within each study at the study level mean. Our simulation studies indicate that AWE is very robust across multiple model violation scenarios we considered, including presence of certain types of nonlinearity in the interaction term.

The rest of the paper 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 the newly proposed method. A comprehensive simulation study was performed to assess the performance of the meta-analysis methods under a variety of scenarios. We primarily focus on the issue of environmental covariate heterogeneity, but also explore several other important factors that could potentially affect the relative performance of these methods: (1) departures from gene–environment independence; (2) heterogeneity in minor allele frequencies (MAFs) across cohorts; (3) lack of a common set of covariates to adjust for across studies; (4) misspecification of the genetic susceptibility model (dominant/codominant/additive); and finally, (5) the presence of a nonlinear form of interaction. In the results section, we report simulation findings followed by an illustrative example, where we examine whether variants 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. This paper is expected to provide useful insights and guidelines for practitioners conducting meta-analyses of GEI.

Methods

Meta-Analysis of GEI Under a Common Effect Model

Consider a quantitative trait Y, a continuous environmental exposure E, a biallelic genetic locus G with genotypes of

AA, *Aa*, and *aa* (where *A* is the minor allele), and other covariates $\mathbf{Z}$. Suppose that there are K independent studies and a total of N participants, with n_k participants in the kth study, $k = 1, \ldots, K$, $\sum_{k=1}^{K} n_k = N$. Let Y_{ki}, E_{ki}, G_{ki}, and $\mathbf{Z}_{ki}$ be the corresponding observations for participant i in study k, for $i = 1, \ldots, n_k$ and $k = 1, \ldots, K$. The *assumed* model for individual responses follows

$$Y_{ki} = \beta_{0k} + \beta_G G_{ki} + \beta_E E_{ki} + \delta G_{ki} E_{ki} + \boldsymbol{\beta}_Z^\top \mathbf{Z}_{ki} + \epsilon_{ki}, \quad (1)$$

where β_{0k} are the study-specific intercepts, β_G, β_E, and $\boldsymbol{\beta}_Z$ are the main effects corresponding to G, E, and $\mathbf{Z}$, and δ is the GEI effect of primary interest. The vector $\boldsymbol{\beta} = (\beta_G, \beta_E, \delta, \boldsymbol{\beta}_Z)$ is assumed to be fixed and common across studies. The random errors ϵ_{ki}'s are assumed as $\epsilon_{ki} \sim N(0, \sigma_k^2)$ and are independent. Our interest lies in estimating the common interaction parameter δ and in testing $H_0 : \delta = 0$.

There are multiple reasons for assuming a common effect model (1). First, this model is used quite frequently in the literature [Hartung et al., 2011; Lin and Zeng, 2010a, 2010b]. Second, the analytical derivation of the relative efficiency and power are facilitated; Third, meta-regression can only be meaningfully conceptualized if the interaction parameter is assumed to be the same across studies; Fourth, with unrelated but distinct fixed effects across studies, it is often hard to find a scientific interpretation of the limiting/expected value of the population parameter to which the standard inverse-variance weighted estimator converges to; thus quantities like bias and mean-squared error become less interpretable. Finally, there was no evidence of effect heterogeneity for the interaction parameters in our T2D data analysis example of eight European studies. However, as we will discuss subsequently, the "common effect" assumption can be relaxed for testing purposes and most of the methods we discuss are valid to test the null hypothesis $H_0 : \delta_1 = \cdots = \delta_K = 0$, if in fact we allowed study specific parameters δ_k in (1).

Various susceptibility models including the 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 codominant model ($G = AA, Aa$ or aa with aa as the reference level) are considered. For codominant model, $\boldsymbol{\beta}_G = (\beta_G^{Aa}, \beta_G^{AA})$ and $\boldsymbol{\delta} = (\delta^{Aa}, \delta^{AA})$ for genotypes Aa and AA can be modified accordingly in model (1).

In the following, we first describe in the existing methods section, three traditional approaches to detect GEI under model (1). The approaches are IPD analysis, standard meta-analysis (UIVW or MIVW) and MR. For the sake of completeness, we also describe a two-step estimator previously suggested by Simmonds and Higgins (2007). We then propose the new AWE in the adaptively weighted estimator section. Throughout the paper, we use the generic notation $\mathbf{v}(\hat{\delta})$ for the asymptotic variance (covariance matrix for multivariate $\hat{\delta}$) of any given estimator $\hat{\delta}$, and $\hat{\mathbf{v}}(\hat{\delta})$ for the corresponding estimated variance.

Existing Methods

(i) Individual patient data analysis (IPD): The IPD analysis fits model (1) using individual level data. Methods such as weighted least square (WLS) can be used to handle the heterogeneous σ_k^2 across studies, if σ_k^2 can be estimated with sufficient accuracy. However, for simplicity, we consider a simple linear regression model that assumes common residual variance $\sigma_k^2 = \sigma^2$ for $k = 1, \ldots, K$ in (1), as the standard implementation of the IPD analysis. Denote the maximum likelihood estimate (MLE) of δ as $\hat{\delta}^{\mathrm{IPD}}$, and its estimated variance as $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{IPD}})$.

(ii) Meta-analysis using inverse-variance weighted estimator: Since the data required for IPD analysis are seldom available in published results, meta-analysis combining summary statistics across individual studies is often used. We consider some variants of IVW estimator under model (1).

(ii.A) UIVW: A UIVW estimator needs the MLEs $\hat{\delta}_k$ and $\hat{\mathbf{v}}(\hat{\delta}_k)$ estimated from model (1) using data from only study k. Under the above model, $\hat{\delta}_k \stackrel{iid}{\sim} N(\delta, \mathbf{v}(\hat{\delta}_k))$ and are independent across studies. The UIVW estimator is given by,

$$\hat{\delta}^{\mathrm{UIVW}} = \left\{ \sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1} \right\}^{-1} \sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1} \hat{\delta}_k$$

$$\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{UIVW}}) = \left\{ \sum_k \hat{\mathbf{v}}(\hat{\delta}_k)^{-1} \right\}^{-1}.$$

The validity of the method requires that $\hat{\delta}_k$ is asymptotically normal $\hat{\delta}_k \stackrel{iid}{\sim} N(\delta, \mathbf{v}(\hat{\delta}_k))$ for a large n_k and the 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 these conditions as "standard conditions" throughout, and we note that it is often implicitly assumed to hold in the traditional meta-analysis literature (e.g. DerSimonian and Laird 1986; Lin and Zeng 2010b; Whitehead and Whitehead 1991).

(ii.B) MIVW: Let $\hat{\boldsymbol{\beta}}_k = (\hat{\beta}_{Gk}, \hat{\beta}_{Ek}, \hat{\delta}_k, \hat{\boldsymbol{\beta}}_{Zk})$ be the MLE of $\boldsymbol{\beta}$ from study k, with estimated multivariate variance-covariance matrix $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)$. 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 Lin and Zeng [2010b],

$$\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}} = \left\{ \sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1} \right\}^{-1} \sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1} \hat{\boldsymbol{\beta}}_k$$

$$\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}}) = \left\{ \sum_k \hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)^{-1} \right\}^{-1}.$$

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 in $\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}}$ and $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}^{\mathrm{MIVW}})$, respectively. Following Lin and Zeng [2010b], $\hat{\delta}^{\mathrm{MIVW}}$ has full asymptotic efficiency as $\hat{\delta}^{\mathrm{IPD}}$ under the common effect model (1). However, the full covariance matrix $\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}_k)$ is not likely to appear in meta-analysis of published results and different studies may adjust for different covariates $\mathbf{Z}$. So $\hat{\delta}^{\mathrm{UIVW}}$ remains the most commonly used meta-analytic

method in spite of potential efficiency loss when compared to $\hat{\delta}^{\mathrm{MIVW}}$ and $\hat{\delta}^{\mathrm{IPD}}$.

(iii) Meta-regression: The assumed model (1) implies that the Y-G association depends linearly on E. So we consider a linear MR model to reveal the underlying dependence between the marginal genetic effects and the aggregated 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

$$Y_{ki} = \lambda_{0k} + \lambda_{Gk}G_{ki} + \boldsymbol{\lambda}_{Zk}^{\top}\mathbf{Z}_{ki} + \eta_{ki}, \quad i = 1, \ldots, n_k. \quad (2)$$

where the errors $\eta_{ki} \sim N(0, \sigma_{\eta k}^2)$, and are independent. At the second step, the MLE $\hat{\lambda}_{Gk}$, the estimated genetic association parameter in study k is regressed on m_k through the MR model

$$\hat{\lambda}_{Gk} = \gamma_0 + \gamma m_k + \nu_k, \quad k = 1, \ldots, K. \quad (3)$$

To account for the potential heterogeneity in $\hat{\mathbf{v}}(\hat{\lambda}_{Gk})$ across studies, we consider the WLS estimator of γ, with weight $w_k = \hat{\mathbf{v}}(\hat{\lambda}_{Gk})^{-1}$ assumed as known, i.e. $\nu_k \sim N(0, \hat{\mathbf{v}}(\hat{\lambda}_{Gk}))$. Denote the WLS estimator of γ in model (3) by $\hat{\delta}^{\mathrm{MR}}$. Then $\hat{\delta}^{\mathrm{MR}}$ and $\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}})$ can be expressed as

$$\hat{\delta}^{\mathrm{MR}} = \frac{\left(\sum_k w_k\right)\left(\sum_k w_k m_k \hat{\lambda}_{Gk}\right) - \left(\sum_k w_k m_k\right)\left(\sum_k w_k \hat{\lambda}_{Gk}\right)}{\left(\sum_k w_k\right)\left(\sum_k w_k m_k^2\right) - \left(\sum_k w_k m_k\right)^2},$$

$$\hat{\mathbf{v}}(\hat{\delta}^{\mathrm{MR}}) = \frac{\sum_k w_k}{\left(\sum_k w_k\right)\left(\sum_k w_k m_k^2\right) - \left(\sum_k w_k m_k\right)^2}\,\hat{\sigma}^2.$$

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 ($\hat{\lambda}_{Gk}$ and $\hat{\mathbf{v}}(\hat{\lambda}_{Gk})$).

(iv) Two-stage estimator: Let $m = \sum_{k,i} E_{ki}/N$ denote the overall sample mean of E, $s_E^2 = N^{-1}\sum_{k,i}(E_{ki} - m)^2$ be the total sample variance of E, $s_{Ek}^2 = n_k^{-1}\sum_{i=1}^{n_k}(E_{ki} - m_k)^2$ be the sample variance of E within the kth study. Denote the corresponding population parameters for m, m_k, s_E^2, s_{Ek}^2 as μ, μ_k, σ_E^2, σ_{Ek}^2, respectively. We make the usual partition 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 paper, we assume $n_k/N \to \varrho_k \in (0, 1)$ as $N \to \infty$. Consider the limiting true population quantities as $tss = \sigma_E^2$, $wss = \sum_k \varrho_k \sigma_{Ek}^2$, and $bss = \sum_k \varrho_k(\mu_k - \mu)^2$. We have $TSS/N \xrightarrow{p} tss$, $WSS/N \xrightarrow{p} wss$, $BSS/N \xrightarrow{p} bss$, as $N \to \infty$.

Motivated by the fact that asymptotic relative efficiency (ARE) of $\hat{\delta}^{\mathrm{MR}}$ compared to $\hat{\delta}^{\mathrm{UIVW}}$ is driven by bss/wss, we consider a two-stage approach analogous to Simmonds and Higgins [2007] as

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

i.e. using $\hat{\delta}^{\mathrm{UIVW}}$ instead of $\hat{\delta}^{\mathrm{MR}}$ if $WSS \geq BSS$ and vice versa. Note that $\hat{\delta}^{\mathrm{TS}}$ is an ad-hoc procedure of discretely determining which method to use, based on the statistic BSS/WSS that measures heterogeneity in E between studies relative to within study variation in E.

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 have significant ecological bias [Berlin et al., 2002; Greenland, 1987; Morgenstern, 1982; Schwartz, 1994] if the aggregate data relationship differs from the one observed in individual level data. Thus, we propose an adaptive estimator that combines $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$ to trade-off between bias and efficiency in a data adaptive way. We first use the following lemma that is also used in Kooperberg and LeBlanc [2008] and Dai et al., [2012].

Lemma 1. *Let Y_i be independent random variables with equal variance, for $i = 1, \ldots, n$, and let $X_j = (X_{1j}, \ldots, X_{nj})^{\top}$ be the data vector corresponding to the jth predictor, $j = 1, \ldots, p + q$. Let $\hat{\lambda}_j$ $(j = 1, \ldots, p)$ and $\hat{\beta}_j$ $(j = 1, \ldots, p + q)$ be the MLEs of the parameters under the two nested linear regression models*

$$Y_i = \lambda_0 + \sum_{j=1}^{p}\lambda_j X_{ij} + \eta_i \quad \text{and} \quad Y_i = \beta_0 + \sum_{j=1}^{p+q}\beta_j X_{ij} + \epsilon_i,$$

then $(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$ and $(\hat{\beta}_{p+1}, \ldots, \hat{\beta}_{p+q})$ are asymptotically independent.

Proof of Lemma 1 is presented in Supplementary Appendix B.1 for the purpose of completeness.

Applying Lemma 1 to models (1) and (2), the marginal genetic association $\hat{\lambda}_{Gk}$ and GEI $\hat{\delta}_k$ are asymptotically independent within each study k, as they are coming from two nested linear regression models using data within study k. 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 $\hat{\lambda}_{Gk}$, then the following corollary holds.

Corollary 1. $\hat{\delta}^{\mathrm{UIVW}}$ *and* $\hat{\delta}^{\mathrm{MR}}$ *are asymptotically independent.*

The independence of the two estimators are critical as we can now borrow the classical idea of constructing an IVW estimator using these two independent ingredients. Assuming the standard conditions, we propose an AWE of the form

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

$$\times \left\{\hat{\mathbf{v}}\left(\hat{\delta}^{\mathrm{UIVW}}\right)^{-1}\hat{\delta}^{\mathrm{UIVW}} + \hat{\mathbf{v}}\left(\hat{\delta}^{\mathrm{MR}}\right)^{-1}\hat{\delta}^{\mathrm{MR}}\right\},$$

which combines $\hat{\delta}^{\mathrm{UIVW}}$ and $\hat{\delta}^{\mathrm{MR}}$ using their inverse-variances as weights. In order to calculate $\hat{\delta}^{\mathrm{AWE}}$, summary statistics of study-specific effect estimates ($\hat{\delta}_k$, $\hat{\mathbf{v}}(\hat{\delta}_k)$, $\hat{\lambda}_{Gk}$, and $\hat{\mathbf{v}}(\hat{\lambda}_{Gk})$) 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.

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

Theorem 1 proved in Supplementary Appendix B.1 establishes the well-known optimality of the inverse variance weights for AWE. A consequence of Theorem 1 is that the precision of $\hat{\delta}^{AWE}$ is the sum of the precisions of $\hat{\delta}^{UIVW}$ and $\hat{\delta}^{MR}$. Under the standard conditions, the estimated variance of the AWE estimator is given by $\hat{\mathbf{v}}(\hat{\delta}^{AWE})^{-1} = \hat{\mathbf{v}}(\hat{\delta}^{UIVW})^{-1} + \hat{\mathbf{v}}(\hat{\delta}^{MR})^{-1}$. We will further show that $\hat{\delta}^{AWE}$ is fully efficient as $\hat{\delta}^{IPD}$ under certain plausible assumptions in the Analytical Results section.

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

Analytical Results

This section presents some analytical results regarding bias, variance, and power properties for the adaptive estimator described in Adaptively Weighted Estimator section. We consider models without covariate $\boldsymbol{Z}$ to simplify the presentation.

Bias

Following classic linear regression and meta-analysis results, $\hat{\delta}^{IPD}$, $\hat{\delta}^{UIVW}$, and $\hat{\delta}^{MIVW}$ are all asymptotically unbiased estimators of δ. However, $\hat{\delta}^{MR}$ 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 is known as "ecological bias" or "ecological fallacy," and is well characterized in the literature [Greenland, 1987; Morgenstern, 1982; Schwartz, 1994]. However, we note that $\hat{\delta}^{MR}$ is an unbiased estimator of δ under the following G-E independence assumption, which is a plausible assumption well-discussed in the literature. We use the generic notation $P(\cdot)$ to denote the distribution of a random variable.

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

Proposition 1. *Under Assumption 1, $\hat{\delta}^{MR}$ as described in model (3) is asymptotically unbiased for δ.*

Proof of Proposition 1 is presented in Supplementary Appendix B.2. In the following Remark 2, we further discuss the issue of potential bias of $\hat{\delta}^{MR}$ and thus in $\hat{\delta}^{AWE}$ (which assigns a positive weight on the MR estimator) if Assumption 1 is violated.

Remark 2 (Bias of $\hat{\delta}^{MR}$ and $\hat{\delta}^{AWE}$). Without Assumption 1, we showed (in Appendix B.2) that the limiting value of the bias of $\hat{\delta}^{MR}$ is proportional to the ratio tss/bss and the correlation between G and E. If the G-E correlations within each study are 0, then $E(\hat{\delta}^{MR}) - \delta \overset{p}{\to} 0$. If Assumption 1 holds, $\hat{\delta}^{AWE}$ is an asymptotically unbiased estimator of δ as both its components are unbiased. Moreover, we show later in variance and relative efficiency section that the limiting value of the weight corresponding to $\hat{\delta}^{MR}$ in $\hat{\delta}^{AWE}$ is bss/tss. So $\hat{\delta}^{AWE}$ adaptively puts less weight on $\hat{\delta}^{MR}$ when the bias of $\hat{\delta}^{MR}$ increases. We find through our numerical investigation that $\hat{\delta}^{AWE}$ is robust to potential ecological bias in $\hat{\delta}^{MR}$ even when the ecological bias is substantial (please see Supplementary Appendix Fig. 9 and Table 6 for the simulation results, and main text Table 4 corresponding to the T2D example).

Variance and Relative Efficiency

Explicit variance formulae $\hat{\mathbf{v}}(\hat{\delta})$ and $\mathbf{v}(\hat{\delta})$ for each estimator of δ are derived under Assumption 1 in Supplementary Appendix B.3. Because the linear regression likelihood $\prod_{k,i} P(Y_{ki}|G_{ki}, E_{ki})$ corresponding to model (1) does not use any assumptions about the joint distribution of G and E, the role of the G-E independence assumption in this paper is only to provide simpler closed form expressions for the variances. This is different from case-control studies where assuming G-E independence and using the retrospective likelihood can lead to large gains in efficiency [Chatterjee and Carroll, 2005; Piegorsch et al., 1994; Umbach and Weinberg, 1997].

In this Variance and Relative Efficiency section, for simplicity of presentation we assume $\sigma_k^2 = \sigma^2$ for $k = 1, \ldots, K$, and consider a dominant susceptibility model for stating Theorems 2 and 3. Let $G = 1$ ($G = 0$) indicate whether an individual is a carrier (noncarrier) of the minor allele A, and let p_k denote $P(G = 1| study = k)$ the carrier frequencies in study k, $k = 1, \ldots, K$.

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

Proof of Theorem 2 is given in Supplementary Appendix B.4. Under Assumption 1, the precision of $\hat{\delta}^{IPD}$ is in general greater than that of $\hat{\delta}^{AWE}$. However, under the additional assumption of homogeneity of the MAFs (Assumption 2 stated below), we have equality $\mathbf{v}(\hat{\delta}^{IPD}) = \mathbf{v}(\hat{\delta}^{AWE})$.

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

Table 1. Glossary of the meta-analysis methods for GEI with summary properties

Methods	Data shared	Bias	ARE[a]	
IPD	individual level data	unbiased	1	
UIVW	$\hat{\delta}_k$, $\hat{v}(\hat{\delta}_k)$	unbiased	wss/tss	under Assumptions 1 and 2
MIVW	$\hat{\beta}_k$, $\hat{v}(\hat{\beta}_k)$	unbiased	1	under assumptions in LZ
MR	$\hat{\lambda}_{Gk}$, $\hat{v}(\hat{\lambda}_{Gk})$ and m_k	unbiased under Assumption 1	bss/tss	under Assumptions 1 and 2
AWE	$\hat{\delta}_k$, $\hat{v}(\hat{\delta}_k)$; $\hat{\lambda}_{Gk}$, $\hat{v}(\hat{\lambda}_{Gk})$ and m_k	unbiased under Assumption 1[b]	1	under Assumptions 1 and 2

[a] ARE, asymptotic relative efficiency as compared to $\hat{\delta}^{IPD}$.
[b] Bias adaptively controlled in AWE.
IPD, individual patient data analysis; UIVW, univariate inverse-variance weighted estimator; MIVW, multivariate inverse-variance weighted estimator; MR, Meta-regression; AWE, adaptively weighted estimator combining UIVW and MR. $\hat{\lambda}_k$ stands for estimates of marginal genetic association and $\hat{\delta}_k$ stands for gene–environment interaction, respectively.

Theorem 3. *Under Assumptions 1 and 2,* $\mathbf{v}(\hat{\delta}^{IPD})^{-1} = \mathbf{v}(\hat{\delta}^{AWE})^{-1} = \mathbf{v}(\hat{\delta}^{UIVW})^{-1} + \mathbf{v}(\hat{\delta}^{MR})^{-1}$, *where* $\mathbf{v}(\hat{\delta}^{UIVW}) = \{Np(1-p)wss\}^{-1}\sigma^2$, $\mathbf{v}(\hat{\delta}^{MR}) = \{Np(1-p)bss\}^{-1}\sigma^2$ *and* $\mathbf{v}(\hat{\delta}^{IPD}) = \mathbf{v}(\hat{\delta}^{AWE}) = \{Np(1-p)tss\}^{-1}\sigma^2$.

Proof of Theorem 3 is given in Supplementary Appendix B.5. Following Theorem 3, the asymptotic variances $\mathbf{v}(\hat{\delta}^{IPD})$, $\mathbf{v}(\hat{\delta}^{UIVW})$, $\mathbf{v}(\hat{\delta}^{MR})$, and $\mathbf{v}(\hat{\delta}^{AWE})$ are all expressed in terms of covariate heterogeneity of E. The ARE between $\hat{\delta}^{UIVW}$ ($\hat{\delta}^{MR}$) and $\hat{\delta}^{IPD}$ is wss/tss (bss/tss). So $\mathbf{v}(\hat{\delta}^{UIVW}) \leq \mathbf{v}(\hat{\delta}^{MR})$, 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. $\mu_k = \mu$), $\mathbf{v}(\hat{\delta}^{UIVW}) = \mathbf{v}(\hat{\delta}^{IPD})$, and $\hat{\delta}^{AWE}$ reduces to $\hat{\delta}^{UIVW}$; in contrast, if all $\sigma^2_{Ek} = 0$ (i.e. E is constant within each study), $\mathbf{v}(\hat{\delta}^{MR}) = \mathbf{v}(\hat{\delta}^{IPD})$, and $\hat{\delta}^{AWE}$ reduces to $\hat{\delta}^{MR}$.

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

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

$\hat{\delta}^{AWE}$ adaptively captures the precision trade-off between the two estimators: $\hat{\delta}^{AWE}$ puts more weight on $\hat{\delta}^{UIVW}$ if WSS is relatively larger than BSS, and vice versa. In summary, under Assumptions 1 and 2, $\hat{\delta}^{AWE}$ is a 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 1. The results in Theorems 2 and 3 are numerically evaluated through a simulation study to examine the effect of relaxing Assumption 1 or 2, and relaxing the homogeneity assumption of σ^2_k.

Remark 3 (Additive and codominant models). In general, it is difficult to provide analytical results related to $\hat{\delta}^{AWE}$ in Theorems 2 and 3 for an additive model, but we can directly translate Theorems 2 and 3 for δ^{Aa} and δ^{AA} respectively under a codominant model if we assume $diag(\hat{\mathbf{v}}(\hat{\lambda}^{Aa}_{Gk}), \hat{\mathbf{v}}(\hat{\lambda}^{AA}_{Gk}))$ for $\hat{\mathbf{v}}(\hat{\lambda}_k)$ in the MR model, i.e. two separate MRs. The statements in Theorems 2 and 3 are numerically evaluated for additive and codominant models through a simulation study relaxing Assumptions 1 or 2, and relaxing the homogeneity assumption of σ^2_k.

Remark 4 (Centering of covariate E). Continuous E is often centered 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 fits model (1) with E centered at their respective study specific means m_k. For the IPD analysis, it is natural to consider that E is centered at the overall mean m. With these centering strategies, $\hat{\delta}^{IPD}$, $\hat{\delta}^{UIVW}$, $\hat{\delta}^{MR}$, and $\hat{\delta}^{AWE}$ remain invariant, and results in Theorems 1–3 still hold for the centered models. The details are shown in Appendix B.4. However, properties of $\hat{\delta}^{MIVW}$ do not hold under the above centering strategy. It is not fully efficient, because m_k-centered model creates artificial heterogeneous main effects of G (depending on m_k) across studies. The m_k-centered model only has two common fixed-effects compared to the true model having three common fixed-effects (Appendix B.4), and this leads to efficiency loss of $\hat{\delta}^{MIVW}$ according to Lin and Zeng [2010b]. In terms of efficiency, $\hat{\delta}^{AWE}$ is preferable to $\hat{\delta}^{MIVW}$ when covariate E is m_k-centered using the study level means.

Remark 5 (Relaxing the common effect assumption). First, we consider the situation with heterogeneous main effects of G and E, namely (β_{Gk}, β_{Ek}) with a common GEI δ across studies in model (1). To handle effect heterogeneity, we could replace (β_G, β_E) by (β_{Gk}, β_{Ek}) in model (1) for the IPD analysis; replace MR model (3) by $\hat{\lambda}_{Gk} = \gamma_{0k} + \gamma m_k + \nu_k$ for the MR analysis; and still use $\hat{\delta}^{UIVW}$ as it does not require homogeneity of (β_{Gk}, β_{Ek}). According to Lin and Zeng [2010b], the modified estimator has the property that $\mathbf{v}(\hat{\delta}^{IPD}) = \mathbf{v}(\hat{\delta}^{MIVW}) = \mathbf{v}(\hat{\delta}^{UIVW})$ because δ is the only common parameter across studies. Theorem 1 still holds since it makes no homogeneity assumption on (β_{Gk}, β_{Ek}). Then we have $\mathbf{v}(\hat{\delta}^{AWE}) < \mathbf{v}(\hat{\delta}^{MIVW})$. In terms of precision, $\hat{\delta}^{AWE}$ is better than $\hat{\delta}^{MIVW}$ when (β_{Gk}, β_{Ek}) are heterogeneous. Next, we consider heterogeneous GEI $\delta_1, \ldots, \delta_K$ across studies in model (1), regardless of (β_{Gk}, β_{Ek}) are homogeneous or not. In this case, it is hard to interpret the expected value of $\hat{\delta}^{UIVW}$, $\hat{\delta}^{MIVW}$, $\hat{\delta}^{MR}$, or $\hat{\delta}^{AWE}$ as a scientifically relevant population parameter. Thus, estimation properties such as bias and mean squared error (MSE) become less meaningful. In this case, we are simply getting an weighted average of within study interaction estimates. Although the analytical results corresponding to $\hat{\delta}^{AWE}$ are derived under a common effect model, the test based on $\hat{\delta}^{AWE}$ is still valid for the null hypothesis $H_0 : \delta_1 = \ldots = \delta_K = 0$. We numerically evaluate the

power and Type-I error under violation of the common effect assumption through simulation 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 $P_w(\delta^*, \alpha) = \Phi(-z_{\frac{\alpha}{2}} + \hat{\mathbf{v}}(\hat{\delta})^{-\frac{1}{2}}\delta^*) + \Phi(-z_{\frac{\alpha}{2}} - \hat{\mathbf{v}}(\hat{\delta})^{-\frac{1}{2}}\delta^*)$, 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 codominant models, we consider a joint Wald test statistic $T = \hat{\delta}^\top \hat{\mathbf{v}}(\hat{\delta})^{-1}\hat{\delta} \overset{H_0}{\sim} \chi_2^2$ for testing $H_0 : \delta = \mathbf{0}$ against $H_1 : \delta \neq \mathbf{0}$, where χ_2^2 is a Chi-square distribution with two degrees of freedom. The power is approximately $P_w(\delta^*, \alpha) = 1 - \Phi_{\chi_2^2}(\chi_{2,\alpha}^2 - \delta^{*\top}\hat{\mathbf{v}}(\hat{\delta})^{-1}\delta^*)$, where $\Phi_{\chi_2^2}$ is the CDF for a χ_2^2 distributed random variable and $\chi_{2,\alpha}^2$ is the corresponding αth upper percentile. The power function $P_w(\delta^*, \alpha)$, or simply P_w, is strictly decreasing in the variance $\hat{\mathbf{v}}(\hat{\delta})$. Thus, the results regarding variances in Theorems 1–3 also determine relative power properties.

Table 1 provides a **glossary table** for all the methods we have discussed, along with their properties, and the summary statistics required to carry out these procedures.

Simulation Study

In order to study the role of G-E independence (Assumption 1) and homogeneity in MAFs across cohorts (Assumption 2), we consider $P(G, E)$ under four different settings: when Assumptions 1 and 2 hold and do not hold. To study the role of covariate heterogeneity in E, we consider cases where *wss* is greater or smaller than *bss*, for a fixed value of *tss*. The details of generating data pair (G_{ki}, E_{ki}) jointly are described in Supplementary Appendix B.6.

Given (G_{ki}, E_{ki}), we then generate the continuous trait Y_{ki} under the IPD model (1), where the study specific intercepts are sampled from $\beta_{0k} \overset{iid}{\sim} U(1.3, 1.5)$, and the true effect sizes $(\beta_E^*, \beta_G^*, \delta^*)$ are determined such that G, E, and GEI explain 1%, 10%, and 0.5–1% of the total variation in Y, respectively, in terms of partial R^2. The residuals follow a $N(0, \sigma_k^2)$ distribution, i.e. no requirement for homogeneity of σ_k^2 is made. In particular, we generate $\sigma_k^2 \overset{iid}{\sim} U(0.3, 0.45)$ that leads to a marginal distribution of $Y \sim N(1.4, 0.4^2)$. The choice of $N(1.4, 0.4^2)$ is motivated by the distribution of log HDL-C level (*mmol/l*) in our T2D dataset. We generate $K = 20$ studies with different sample sizes involving a total of $N = 10,000$ participants ($n_k = 200$, for $k = 1, \ldots, 6$; $n_k = 400$, for $k = 7, \ldots, 11$; $n_k = 500$, for $k = 12, \ldots, 17$; $n_{18} = 800$; $n_{19} = 1000$; $n_{20} = 2000$).

We calculate $\hat{\delta}$ and $\hat{\mathbf{v}}(\hat{\delta})$ corresponding to each proposed estimator, including $\hat{\delta}^{\text{IPD}}$, $\hat{\delta}^{\text{UIVW}}$, $\hat{\delta}^{\text{MIVW}}$, $\hat{\delta}^{\text{MR}}$, $\hat{\delta}^{\text{TS}}$, and $\hat{\delta}^{\text{AWE}}$. We carry out $R = 1,000$ replications under each setting, and summarize the results in terms of relative bias $(\frac{1}{R}\sum_{r=1}^{R}\hat{\delta}_{(r)} - \delta^*)/\delta^* \times 100\%$, average model-based variance $\frac{1}{R}\sum_{r=1}^{R}\hat{\mathbf{v}}(\hat{\delta}_{(r)})$, empirical variance $\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\delta}_{(r)} - \bar{\hat{\delta}}_{(r)})^2$, MSE

$(\frac{1}{R}\sum_{r=1}^{R}(\hat{\delta}_{(r)} - \delta^*)^2)$, power (proportion of simulations rejecting the null hypothesis using the Wald test) and Type-I error (proportion of simulations rejecting the null hypothesis when the data are generated under the null).

Lack of common set of covariates across studies: We then consider covariate $\mathbf{Z} = (Z^1, Z^2, Z^3)$ that stand for typical covariates (age, sex, race) in the IPD model (1). In particular, we consider a hypothetical study where 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), with $\boldsymbol{\beta}_Z^*$ is determined such that the Type-III partial R^2 corresponding to (Z^1, Z^2, Z^3) is (2%, 1%, 1%), respectively. Let $\mathbf{Z}_k$ be the set of covariates for the kth study. We consider an analysis where $\mathbf{Z}_k$ is only partially available from individual studies, and refer to this situation as "lack of common set of covariates across studies." In particular, we consider $\mathbf{Z}_k = (Z_k^1, Z_k^2, Z_k^3)$ for $k = 1, 2, 3$; $\mathbf{Z}_k = (Z_k^1, Z_k^2)$ for $k = 4, 5, 6$; $\mathbf{Z}_k = (Z_k^1, Z_k^3)$ for $k = 7, 8, 9$; $\mathbf{Z}_k = (Z_k^2, Z_k^3)$ for $k = 10, 11, 12$; $\mathbf{Z}_k = Z_k^1$ for $k = 13, 14$; $\mathbf{Z}_k = Z_k^2$ for $k = 15, 16$; $\mathbf{Z}_k = Z_k^3$ for $k = 17, 18$; No $\mathbf{Z}_k$ available 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 the above setting. We refer it as a naive IPD estimator ($\hat{\delta}^{\text{NIPD}}$). For the meta-analysis, we obtain $\hat{\delta}^{\text{UIVW}}$ and $\hat{\delta}^{\text{MIVW}}$ from the kth study model adjusted for available $\mathbf{Z}_k$, for $k = 1, \ldots, K$. For MR, we adjust for $\mathbf{Z}_k$ at the first stage in the marginal genetic association model, and regress the MLEs of adjusted effects of G on m_k. These methods are compared with an ideal IPD estimator $\hat{\delta}^{\text{IPD}}$ that adjusted for all $\mathbf{Z}$. This ideal IPD estimator can only be computed in the simulation study but can not be calculated in reality with some studies missing components of $\mathbf{Z}$.

Nonlinear GEI model: We consider a nonlinear GEI model where the phenotype-genotype association parameter $\beta_G(E)$ varies with E through a sigmoid function $\beta_G(E) = 2\exp(E - 50)/\{1 + \exp(E - 50)\} + 2$, instead of the assumed linear interaction (shown in Fig. 1). In this case, $\beta_G(E)$ changes at different rates for different values of E (sharper around the mean value of E, flatter at more extreme values of E), which leads to nonlinear interaction. In Figure 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 the true nonlinear interaction. In the simulation study, we generate $K = 20$ studies, where four studies have relatively larger within study variability (studies 5, 10, 11, 15 in Figs. 1 and 2) as compared to the other 16 studies. The complete description of n_k, μ_k, and σ_{Ek} for the 20 studies are given in Figures 1 and 2. We generate Y through the nonlinear interaction model $Y_{ki} = \beta_{0k} + \beta_G(E_{ki})G_{ki} + \varepsilon_{ki}$, where $\varepsilon_{ki} \overset{iid}{\sim} N(0, \sigma_k^2)$ and are independent. The within study relationships of the marginal effects of G as a function of E, namely, $\beta_G(E)$ are substantially different across studies. The effect heterogeneity and nonlinearity might influence the validity and relative

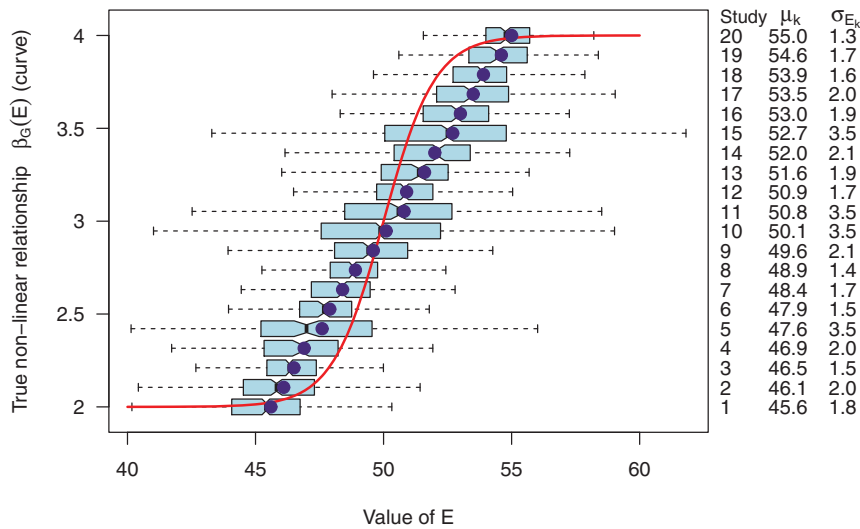## Covariate heterogeneity of E across studies (boxplots)

Study	μ_k	σ_{E_k}
20	55.0	1.3
19	54.6	1.7
18	53.9	1.6
17	53.5	2.0
16	53.0	1.9
15	52.7	3.5
14	52.0	2.1
13	51.6	1.9
12	50.9	1.7
11	50.8	3.5
10	50.1	3.5
9	49.6	2.1
8	48.9	1.4
7	48.4	1.7
6	47.9	1.5
5	47.6	3.5
4	46.9	2.0
3	46.5	1.5
2	46.1	2.0
1	45.6	1.8

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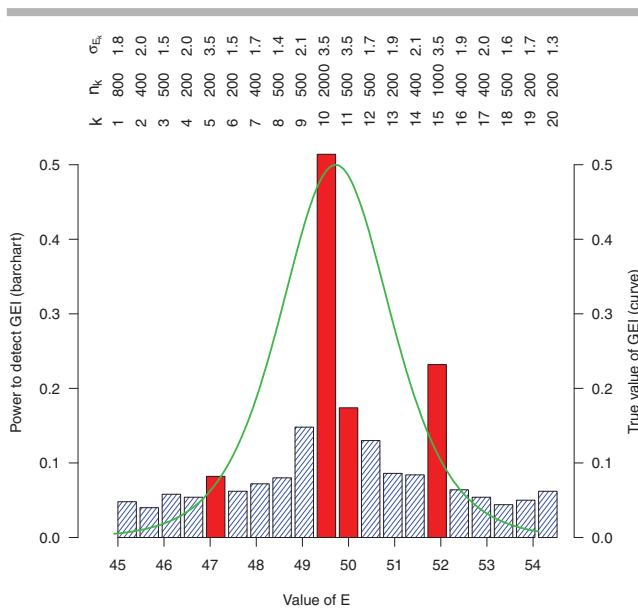

Figure 2. Nonlinear GEI model: the height of the bars represent the power to detect GEI across individual studies; the (green) curve shows the value of the true nonlinear GEI parameter; the top panel shows the sample sizes n_k and the within study standard deviations σ_{E_k} of E, the four studies with relatively greater σ_{E_k} are highlighted (in red).

performance of the proposed methods where a linear form of interaction is assumed. Therefore, we evaluate the robustness of the proposed meta-analysis estimators under this nonlinear GEI model.

Results

Simulation Results

The simulation results are numerically presented in terms of bias, variance, MSE, and power (Supplementary Appendix Tables 1–4). The relative performances of the methods are very similar across all three susceptibility models and all four settings, so we only present in the main text the most general setting where the data are generated without either Assumption 1 or 2. The detailed simulation results are given in the Appendix, and we only summarize the key features in the following.

Gene–environment independence and ecological bias: For bias, comparing settings with and without G-E independence, we observe no substantial difference among the proposed estimators, including the potentially biased estimators $\hat{\delta}^{MR}$ and $\hat{\delta}^{AWE}$ (Supplementary Appendix Tables 1–4). When Assumption 1 is relaxed but G-E correlation is not very strong (Appendix Tables 3 and 4), the magnitude of relative bias of $\hat{\delta}^{MR}$ may be up to $\pm 7\%$ but bias of $\hat{\delta}^{AWE}$ is still well controlled (up to $\pm 4\%$). Comparing to the Monte Carlo error (up to $\pm 3\%$ even for the unbiased estimators), the bias of $\hat{\delta}^{AWE}$ is not to a level of practical concern even when there is some bias in $\hat{\delta}^{MR}$. In additional simulation settings where $\hat{\delta}^{MR}$ is susceptible to substantial ecological bias (up to 35%) and when G-E correlation is extremely strong, our results (Supplementary Appendix Fig. 9 and Table 6) indicate the adaptive feature of $\hat{\delta}^{AWE}$ in controlling the bias from $\hat{\delta}^{MR}$ by assigning decreased weight. The relative bias of $\hat{\delta}^{AWE}$ is still < 5%. Thus the issue of ecological bias for aggregate analysis in $\hat{\delta}^{MR}$ is less of a concern for $\hat{\delta}^{AWE}$. For variance, we did not observe precision gain by making the G-E

independence assumption as expected. Results stated in Theorem 2 appear to hold numerically for all three genetic susceptibility models, even when Assumption 1 is relaxed (Supplementary Appendix Tables 3 and 4).

Homogeneity in allele frequencies across cohorts: Comparing settings with and without homogeneous allele frequencies across studies, we did not observe any appreciable differences in results. Results in Theorem 3 appear to hold numerically for all three genetic susceptibility models, even when Assumption 2 is relaxed.

Covariate heterogeneity in E: We observe that the ARE between $\hat{\delta}^{\text{UIVW}}$ ($\hat{\delta}^{\text{MR}}$) and $\hat{\delta}^{\text{IPD}}$ can be well characterized in terms of wss/tss (bss/tss), respectively. We found that $\hat{\delta}^{\text{UIVW}}$ is more efficient than $\hat{\delta}^{\text{MR}}$ if $wss > bss$, and vice versa. The precision trade-off is captured well by the adaptively determined weights in $\hat{\delta}^{\text{AWE}}$. We observe that $\hat{\delta}^{\text{AWE}}$ is more efficient than the usual meta-analytic estimators $\hat{\delta}^{\text{UIVW}}$, $\hat{\delta}^{\text{MR}}$, or $\hat{\delta}^{\text{TS}}$, and had almost the same efficiency as $\hat{\delta}^{\text{IPD}}$ and $\hat{\delta}^{\text{MIVW}}$ under all simulation scenarios. The findings with finite samples are consistent with our analytical results in Theorems 2 and 3 and Lin and Zeng [2010b].

In terms of power, we find the proposed methods (IPD, UIVW, MIVW, MR, TS, and AWE) are divided into three groups in Figure 3, as expected. Group 1: IPD, MIVW, and AWE; group 2: UIVW; group 3: MR. Group 1 has the most powerful tests, which is consistent with our analytical results and Lin and Zeng [2010b]; group 2 is more powerful than group 3 if $bss < wss$, and vice versa. TS performs similarly as the better group between groups 2 and 3. The empirical estimates of Type-I error are close to the true 0.05 level for all tests.

Heterogeneous GEI effects across studies: We examine the power and Type-I error corresponding to the proposed methods, where $\delta_1, \ldots, \delta_K$ are heterogeneous across studies in model (1). In particular, for each true effect size δ^* determined by a given R^2 as discussed in the Simulation Study section, we generate $\delta_k^* \overset{iid}{\sim} U(0, 2\delta^*)$ that vary across studies but has the same mean δ^*. Supplementary Appendix Table 5 shows the power and Type-I error under data generated from the heterogeneous GEI model, where we do not observe any substantial difference with the testing results from homogeneous GEI model.

Misspecification of the genetic susceptibility model: We examine the power under misspecified susceptibility models (dominant/additive), where the true generating model is codominant. 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 the codominant model (shown in Supplementary Appendix Fig. 7), because the misspecification is not strong and the fitted dominant or additive models use 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 codominant model (shown in Fig. 5). Thus, it could happen that

Table 2. Comparison of methods in terms of estimate, standard error of the estimate and power for GEI, under a simulation study of nonlinear GEI

Methods[a]	Estimate	SE[b]	Power
IPD	0.21	0.045	0.98
UIVW	0.18	0.070	0.69
MIVW	0.21	0.045	0.98
MR	0.23	0.060	0.82
AWE	0.21	0.045	0.98
TS			0.85

[a] IPD, individual patient data; UIVW, univariate inverse-variance weighted estimator; MIVW, multivariate inverse-variance weighted estimator; MR, Meta-regression; AWE, adaptively weighted estimator combining UIVW and MR; TS, two-stage approach.
[b] SE: standard error.

the codominant model has more power compared to other simpler models, though it uses additional parameters for capturing GEI.

Lack of common set of covariates across studies: Figure 4 shows the power curves under this situation without either Assumption 1 or 2. Compared to the basic setting without covariate adjustment (Fig. 3), there is no substantial difference in the relative performances of these methods. We observe that the GEI estimate $\hat{\delta}$ and variance $\hat{v}(\hat{\delta})$ is fairly stable, though the main effects of $\hat{\beta}_G$ and $\hat{\beta}_E$ are substantially influenced under this situation. VanderWeele et al., [2012] also showed similar results that, under G-E 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 various other settings with similar results are given in Supplementary Appendix Figures 4–6.

Nonlinear GEI model: When the IPD are generated under the nonlinear GEI model described in Figures 1 and 2, the power to detect GEI from individual studies is very low (< 0.25), except study 10 where the sample size n_{10}, effect size (depends on E) and variance $\sigma_{E_{10}}^2$ are all relatively larger than the other studies (Fig. 2). In Table 2, IPD, MIVW, and AWE show the highest powers. Because most of the 20 studies are unable to capture the true nonlinear GEI, especially those with restricted range of E at the two extremes of the E distribution, the nonlinearity of GEI leads to the low power of $\hat{\delta}^{\text{UIVW}}$. In this particular example, we observe that power for meta-regression is higher than power for UIVW meta-analysis. Instead of choosing alternatively between $\hat{\delta}^{\text{UIVW}}$ and $\hat{\delta}^{\text{MR}}$, we can use $\hat{\delta}^{\text{AWE}}$ as the default meta-analytic estimator. The relative performance of $\hat{\delta}^{\text{AWE}}$ is close to $\hat{\delta}^{\text{IPD}}$. 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 certain types of model misspecification that univariate meta-analysis methods can not.

Data Analysis for a Set of Studies Investigating T2D

The proposed methods are applied to a set of studies investigating T2D, including eight European cohorts: FIN-D2D
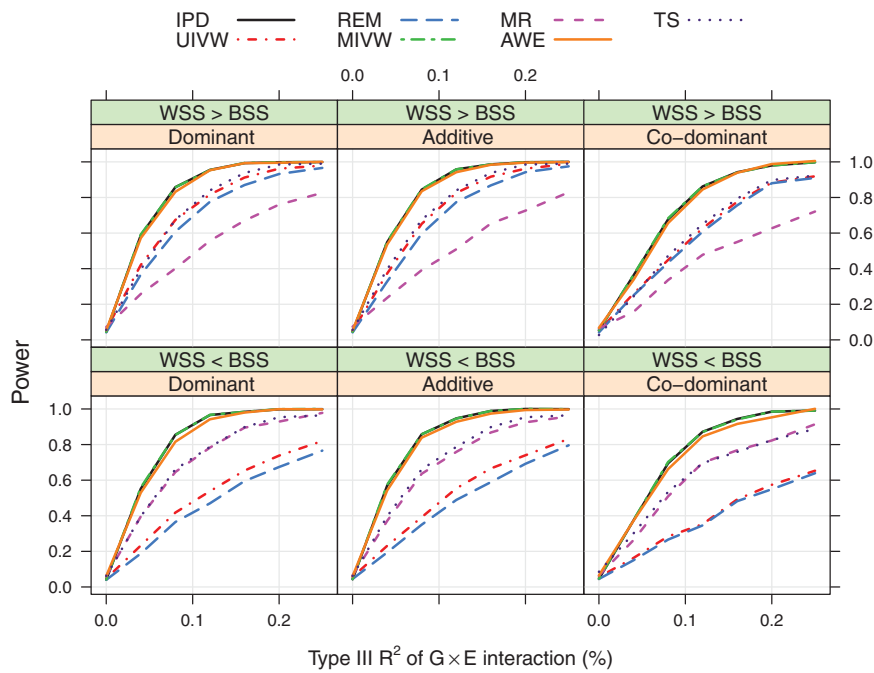
Figure 3. Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study, where data are simulated without any assumption on gene–environment independence or homogeneity in allele frequencies across studies.
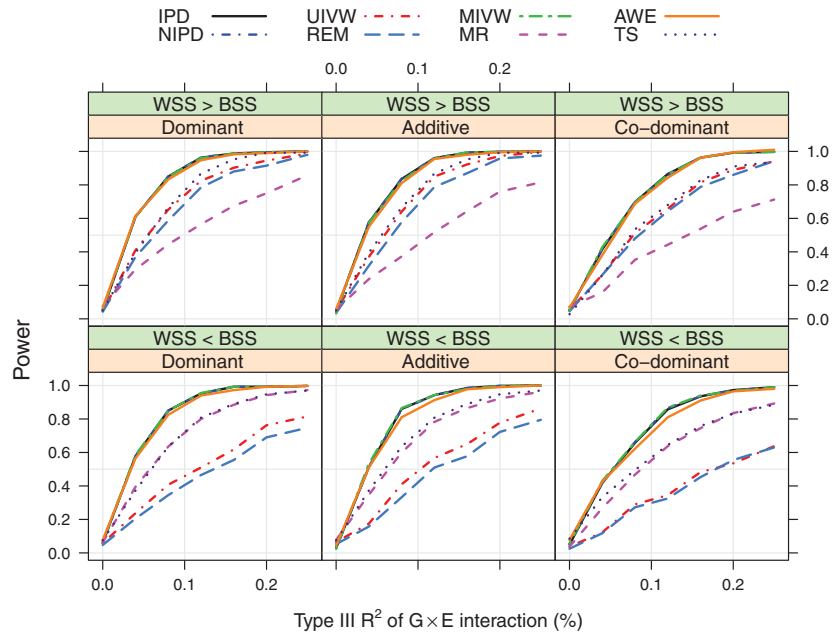

Figure 4. Comparison of the proposed meta-analytical methods (in terms of power) under different scenarios of susceptibility models and covariate heterogeneity through a simulation study (representing the situation of lack of common set of covariates across studies), where data are simulated without any assumption on gene–environment independence or homogeneity in allele frequencies.
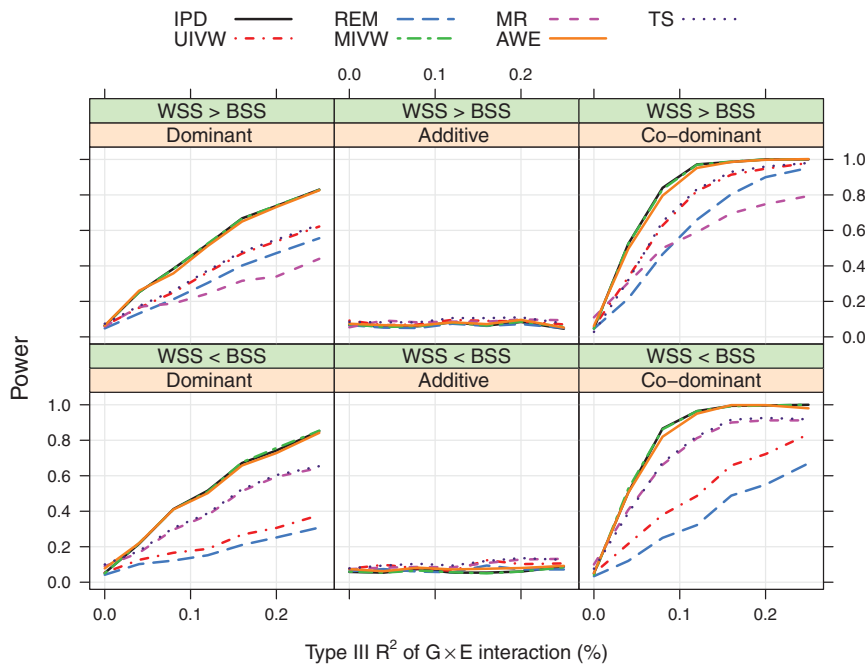
Figure 5. Power curves under misspecified susceptibility models (dominant/additive), where the generating codominant model has $\delta^{AA} = -\delta^{Aa}$, where data are simulated without any assumption on gene–environment independence or homogeneity in allele frequencies across studies.

2007 study (D2D2007), DIAbetes GENetic study (DIAGEN), Finnish Diabetes Prevention Study (DPS), Finland-United States Investigation of NIDDM Genetics study (FUSION, FUSION S2), Nord-Trϕndelag Health Study 2 (HUNT), METabolic Syndrome in Men study (METSIM), and Tromsϕ study (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 [Voight et al., 2010; Zeggini et al., 2008], where the variants at *FTO* gene are known to influence T2D predisposition through an effect on BMI. Age, BMI, and sex are all known risk factors for T2D and the T2D-related quantitative trait HDL-C [Morris et al., 2012; Scott et al., 2012]. In this paper, we investigated whether SNPs in *FTO* gene modifies the effect of environmental factors (e.g. age and BMI) on HDL-C. The effect modification characterized by interaction on HDL-C

has not been reported so far, though marginal association between SNPs in *FTO* and HDL-C have been noted previously [Doney et al., 2009; Kring et al., 2008].

Among the eight cohorts, we have a total of $N = 11,151$ genotyped participants who have HDL-C levels, age, sex, and BMI available, with sample sizes ranging between 172 and 2,729. Participants known to be on lipid medication are excluded from this analysis. The descriptive summary statistics for the eight cohorts are shown in Table 3. Since the SNPs we initially examined (namely, 10 SNPs in *FTO* strongly associated with T2D/obesity/BMI that are listed on the National Human Genome Research Institute GWAS catalog) are in high linkage disequilibrium and show very similar results, we only present our results for one representative SNP, rs1121980. The SNP's genotype distribution did not significantly deviate from that expected under

Table 3.   Summary statistics for the eight European cohorts

Cohorts[a]	N	T2D (%)	HDL-C (*mmol/l*) Mean (SD)	rs1121980 MAF[b]	Age (year) Mean (SD[b])	BMI (*kg/m²*) Mean (SD)	Gender Female (%)	SNP∗Age Corr (P)[c]	SNP∗BMI Corr (P)
D2D2007	2,116	14	1.46 (0.35)	0.41	58.8 (8.3)	27.2 (4.8)	54	−0.03 (0.19)	0.03 (0.24)
DIAGEN	1,510	29	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	0.0	1.22 (0.29)	0.44	55.1 (7.1)	31.3 (4.6)	68	−0.02 (0.69)	0.16 (<.01)
FUSION	172	0.0	1.29 (0.32)	0.43	38.6 (10.9)	26.2 (4.9)	55	0.04 (0.56)	0.23 (<.01)
FUSION-S2	2,729	31	1.45 (0.41)	0.40	57.3 (8.4)	27.9 (5.1)	44	−0.02 (0.22)	0.06 (<.01)
HUNT	1,324	43	1.26 (0.38)	0.47	67.2 (13.1)	28.0 (4.4)	48	<.01 (0.94)	0.06 (0.03)
METSIM	1,456	43	1.42 (0.40)	0.44	56.3 (6.6)	27.9 (4.7)	0	−0.05 (0.08)	0.03 (0.32)
TROMSO	1,411	50	1.43 (0.42)	0.49	59.9 (12.5)	27.6 (4.7)	50	<.01 (0.91)	0.04 (0.15)
Entire study	11,151	31	1.41 (0.40)	0.44	59.4 (11.3)	27.8 (4.9)	44	<.01 (0.85)	0.05 (<.01)

[a] Data reflect patients who were genotyped from the eight European cohorts.
[b] SD, standard deviation; BMI, body mass index; MAF, minor allele frequency.
[c] Corr(P), Spearman correlation between SNP rs1121980 and environmental factor with corresponding *P*-value.

Table 4. IPD/Meta-analysis results of GEI for the T2D study, where log transformed HDL-C level was regressed on SNP, age, BMI, sex, T2D status, cohorts, and SNP × E interaction (E as BMI and age in two separate analysis) in the IPD model. Estimates, SEs and CIs have been multiplied by 1,000

| Methods[a] | rs1121980 (additive) × BMI | | | P-value* | |
	Estimate	SE[b]	95% CI[b]	Additive	Codominant
IPD	1.474	0.687	(0.128, 2.821)	0.03[c,d]	0.03[c]
UIVW	1.731	0.675	(0.407, 3.054)	0.01[c]	0.02[c]
MIVW	1.518	0.663	(0.219, 2.816)	0.02[c]	0.01[c]
MR	−0.719	3.136	(−6.866, 5.429)	0.81	0.69
AWE	1.622	0.660	(0.328, 2.916)	0.01[c]	0.02[c]
	rs1121980 (additive) × age			Additive	Codominant
IPD	0.011	0.304	(−0.585, 0.606)	0.97	0.68[d]
UIVW	0.046	0.337	(−0.613, 0.706)	0.89	0.74
MIVW	−0.008	0.307	(−0.610, 0.594)	0.97	0.69
MR	0.180	0.522	(−0.843, 1.203)	0.73	0.77
AWE	0.086	0.283	(−0.469, 0.640)	0.76	0.69

[a] IPD, individual patient data; UIVW, univariate inverse-variance weighted estimator; MIVW, multivariate inverse-variance weighted estimator; MR, Meta-regression; AWE, adaptively weighted estimator combining UIVW and MR.
[b] SE, standard error; CI, confidence interval.
[c] Indicating significance at α = 0.05 level.
[d] Indicating whether additive or codominant model has smaller AIC under the IPD model.

Hardy–Weinberg equilibrium, and we did not need any imputation given the missing genotype proportion < 0.1%. The MAF of rs1121980 ranges from 0.40–0.49 across cohorts, as a suggestive evidence for no violation of Assumption 2. As in Table 3, the mean age ranges from 55–67 years except FUSION (mean age = 39). This cohort is younger because it is actually a subcohort of the original FUSION study, with either spouse or offspring of selected subjects from the original study. The mean BMI ranges from 26–28 kg/m^2 except the DPS cohort (mean BMI = 31). DPS cohort has an inclusion criterion of having BMI> 25 at baseline. The covariate heterogeneity of E between cohorts is relatively small, where BSS_{age}/TSS_{age} = 15% and BSS_{BMI}/TSS_{BMI} = 2%, respectively. The two "outlier" cohorts, FUSION and DPS, both have only very small sample sizes compared to the other studies, so their influence on UIVW and MIVW is expected to be small. However, their influence on MR could be substantial due to a small number of studies.

Analysis Model: The IPD model we fitted is given by

$$\log(\text{HDL-C}_{ki}) = \beta_{0k} + \beta_G G_{ki} + \delta G_{ki} \times E_{ki} + \beta_a \text{age}_{ki}$$
$$+ \beta_b \text{BMI}_{ki} + \beta_s \text{sex}_{ki} + \beta_t \text{T2D}_{ki} + \epsilon_{ki}, \quad (4)$$

for $k = 1, \ldots, 8$; $i = 1, \ldots, n_k$. In model (4), SNP rs1121980 is used for G with both additive and codominant coding; BMI and age is used as E in two separate analyses; T2D status is adjusted to account for biased sampling of the genotyped subjects (more T2D cases are genotyped than noncases). HDL-C is log-transformed in order to reduce the skewness of its distribution. The proposed methods, including IPD, UIVW, MIVW, MR, and AWE, are implemented and compared. G-E independence appears to be violated for rs1121980×BMI analysis (Spearman correlations across studies are reported

in Table 3). This is expected as FTO is an obesity related gene. G-E independence does not appear to be violated for rs1121980×age (Table 3).

Results: Figure 6 shows the forest plots of estimated GEI from individual cohorts and the combined estimates using joint analysis and meta-analysis. The corresponding numerical results are summarized in Table 4. There is no evidence of effect heterogeneity for both rs1121980×BMI (P = 0.87) and rs1121980×age (P = 0.90) interactions based on Cochran's Q test, so we proceed with a common effect model. In Figure 6 and Table 4, all these meta-analytical methods UIVW, MIVW, and AWE showed very similar results as IPD, except MR. For example, for studying the interaction rs1121980×BMI, the marginal SNP effects of rs1121980 against mean BMI values across cohorts are plotted in Supplementary Appendix Figure 8, where MR is very sensitive to the outliers as the number of cohorts is small (K = 8). MR also appears to lack efficiency due to small K and the small ratio BSS/WSS. However, $\hat{\delta}^{AWE}$ is robust to the bias from $\hat{\delta}^{MR}$ since it only assigned a weight of 0.05 on $\hat{\delta}^{MR}$ ($\hat{\mathbf{v}}(\hat{\delta}^{MR})^{-1}/\{\hat{\mathbf{v}}(\hat{\delta}^{UIVW})^{-1} + \hat{\mathbf{v}}(\hat{\delta}^{MR})^{-1}\}$). This is further evidence that $\hat{\delta}^{AWE}$ can data adaptively shrink to the "better" estimator.

In the interaction model (4), positive rs1121980×BMI interactions are found under all proposed methods (except MR) in Table 4, with P-values ranging from 0.01 to 0.03 for the additive model. In particular, the estimates obtained from model (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, HDL-C level on average decreased by 1.53% (95% CI: (1.37, 1.70)) given rs1121980 = GG, by 1.39% (95% CI: (1.28, 1.49)) given rs1121980 = AG or GA; and by 1.24% (95% CI: (1.06, 1.42)) given rs1121980 = AA; (2) under codominant model, HDL-C level decreased by 1.51% (95% CI: (1.33, 1.69)) given rs1121980 = GG, by 1.41% (95% CI: (1.27, 1.56)) given rs1121980 = AG or GA; and by 1.21% (95% CI: (0.99, 1.42)) given rs1121980 = AA. The results under additive and codominant models are 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 low-density lipoprotein cholesterol (LDL-C), total cholesterol or LDL-C/HDL-C ratio. The suggestive effect modification of BMI by the SNPs on FTO that we have found for HDL-C needs to be replicated in independent studies and validated in larger meta-analysis.

Discussion

In this paper, we proposed and compared a set of meta-analysis approaches for analyzing GEI with quantitative traits. We showed the proposed AWE, as a combination of meta-analysis and meta-regression estimators, performed better than discretely 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 environmental covariate heterogeneity through the ratio of
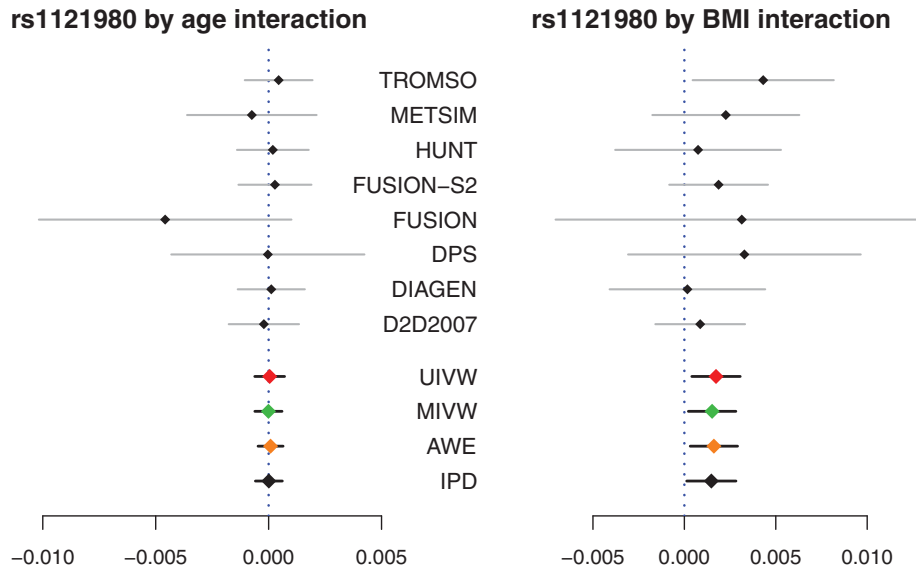
rs1121980 by age interaction

rs1121980 by BMI interaction

Figure 6. Forest plots showing the estimated gene–environment interactions (under additive model of rs1121980) across the eight European cohorts, as well as the combined estimates through meta-analysis. [IPD, individual patient data; UIVW, univariate inverse-variance weighted estimator; MIVW, multivariate inverse-variance weighted estimator; AWE, adaptively weighted estimator combining UIVW and meta-regression.]

the between and within study variances of the covariate E, and, that the AWE adaptively weights its component estimators to minimize the variance of the resulting hybrid estimator. The resulting AWE retains full efficiency of the joint analysis using IPD under certain assumptions. The AWE is very simple to calculate based on summary statistics for marginal genetic association and gene–environment interaction parameters (estimate and standard error) along with the covariate mean of E (see Table 1). The computation is simple and scalable to genome-wide analysis. We suggest possible use of 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. The features included: (1) departures from G-E independence; (2) heterogeneity in MAFs across cohorts; (3) lack of a common set of covariates across studies; (4) misspecification of the genetic susceptibility model (dominant/codominant/additive); and (5) the presence of a nonlinear form of interaction. Under all the above situations, we found the performance of AWE is close to IPD estimator. In particular, under the nonlinear interaction model setting, where standard meta-analytical technique failed and the AWE is able to capture the lost efficiency based on the summary data. We also reported some suggestive evidence for GEI between rs1121980 on the *FTO* gene and BMI on HDL-C levels.

As a reviewer has pointed out, we are risking some bias for gaining precision in AWE by including MR as a component, and MR is susceptible to ecological bias. However, as we note in the analysis of the T2D example in Table 4, where the MR estimate is quite different from the rest, the AWE aligns itself with the more sensible UIVW estimator. Our simulation results also indicate this adaptive feature of AWE in controlling the bias from the MR component by assigning

decreased weight to it (Supplementary Appendix Fig. 9 and Table 6). Moreover, regardless of ecological fallacy, under the additional assumption of G-E independence, AWE remains unbiased. Thus the issue of ecological bias for aggregate analysis in MR is less of a concern for AWE. We also noted that AWE performs well across the whole spectrum of BSS/TSS ratio, not just intermediate values of this quantity (Supplementary Appendix Fig. 9 and Table 6).

We have mainly focused on quantitative traits with an underlying common fixed effect model. The potential limitation of this approach is that the results might not translate directly to dichotomous traits under a case-control design, where assuming G-E independence can lead to huge gain in efficiency [Chatterjee and Carroll, 2005; Piegorsch et al., 1994; Umbach and Weinberg, 1997]. 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. Sample code for all methods is available at http://www-personal.umich.edu/ bhramar/software/.

References

Aschard H, Hancock DB, London SJ, Kraft P. 2011. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Hum Hered* 70:292–300.

Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study Group. 2002. Individual patient-versus group-level data

meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 21:371–387.

Borenstein M, Hedges LV, Higgins J, Rothstein HR. 2010. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synthesis Methods* 1:97–111.

Chatterjee N, Carroll RJ. 2005. Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* 92:399–418.

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

Dai J, Kooperberg C, Leblanc M, Prentice RL. 2012. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika* 99:929–944.

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. *Controlled Clin Trials* 7:177–188.

Doney A, Dannfald J, Kimber C, Donnelly LA, Pearson E, Morris AD, Palmer CN. 2009. The FTO gene is associated with an atherogenic lipid profile and myocardial infarction in patients with type 2 diabetes: a genetics of diabetes audit and research study in Tayside Scotland (Go-DARTS) study. *Circ Cardiovasc Genet* 2:255–259.

Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, Wheeler E, Glazer NL, Bouatia-Naji N, Gloyn AL and others. 2010. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 42:105–116.

Fleiss J. 1993. Review papers: the statistical basis of meta-analysis. *Stat Methods Med Res* 2:121–145.

Greenland S. 1987. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 9:1–30.

Hartung J, Knapp G, Sinha BK. 2011. *Statistical Meta-analysis with Applications*. Wiley, New York.

Khoury M, Wacholder S. 2009. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies: challenges and opportunities. *Am J Epidemiol* 169:227–230.

Kilpeläinen T, Qi L, Brage S, Sharp SJ, Sonestedt E, Demerath E, Ahmad T, Mora S, Kaakinen M, Sandholt CH and others. 2011. Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218,166 adults and 19,268 children. PLoS Med 8:e1001116.

Kooperberg C, LeBlanc M. 2008. Increasing the power of identifying gene×gene interactions in genome-wide association studies. *Genet Epidemiol* 32:255–263.

Kovalchik SA. 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.

Kring S, Holst C, Zimmermann E, Jess T, Berentzen T, Toubro S, Hansen T, Astrup A, Pedersen O, Sørensen TI. 2008. FTO gene associated fatness in relation to body fat distribution and metabolic traits throughout a broad range of fatness. PLoS One 3:e2958.

Lin D, Zeng D. 2010a. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol* 34:60–66.

Lin D, Zeng D. 2010b. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* 97:321–332.

Manning A, LaValley M, Liu C, Rice K, An P, Liu Y, Miljkovic I, Rasmussen-Torvik L, Harris TB, Province MA and others. 2011. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP×environment regression coefficients. *Genet Epidemiol* 35:11–18.

Morgenstern H. 1982. Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 72:1336–1344.

Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè AV, Steinthorsdottir V, Strawbridge RJ, Khan H, Grallert H, Mahajan A and others. 2012. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* 44:981–990.

Mukherjee B, Ahn J, Gruber S, Chatterjee N. 2012. Testing gene-environment interaction in large-scale case-control association studies: possible choices and comparisons. *Am J Epidemiol* 175:177–190.

Piegorsch WW, Weinberg CR, Taylor JA. 1994. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med* 13:153–162.

Psaty B, O'Donnell C, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, Uitterlinden AG, Harris TB, Witteman JC, Boerwinkle E and others. 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. *Circ Cardiovasc Genet* 2:73–80.

Sarwar N, Butterworth A, Freitag D, Gregson J, Willeit P, Gorman DN, Gao P, Saleheen D, Rendon A, Nelson CP and others. 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 LF, Garavito ML, Braun T, Bjonnes A, Young R, Ho WK, Rasheed A, Frossard P and others. 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. *Am J Public Health* 84:819–824.

Scott L, Mohlke K, Bonnycastle L, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU and others. 2007. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316:1341–1345.

Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, Mägi R, Strawbridge RJ, Rehnberg E, Gustafsson S and others. 2012. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet* 44:991–1005.

Simmonds M, Higgins J. 2007. Covariate heterogeneity in meta-analysis: criteria for deciding between meta-regression and individual patient data. *Stat Med* 26:2982–2999.

Song C, Chen GK, Millikan RC, Ambrosone CB, John EM, Bernstein L, Zheng W, Hu JJ, Ziegler RG, Nyante S and others. 2013. A genome-wide scan for breast cancer risk haplotypes among African American women. *PloS One* 8:e57298.

Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen H, Lindgren CM, Luan J, Mägi R and others. 2010. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42:937–948.

Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ and others. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nat Genet* 466:707–713.

Umbach DM, Weinberg CR. 1997. Designing and analysing case-control studies to exploit independence of genotype and exposure. *Stat Med* 16:1731–1743.

VanderWeele T, Mukherjee B, Chen J. 2012. Sensitivity analysis for interactions under unmeasured confounding. *Stat Med* 31:2552–2564.

Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C, Welch RP, Zeggini E, Huth C, Aulchenko YS, Thorleifsson G and others. 2010. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 42:579–589.

Whitehead A, Whitehead J. 1991. A general parametric approach to the meta-analysis of randomized clinical trials. *Stat Med* 10:1665–1677.

Willer C, Li Y, Abecasis G. 2010. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 26:2190–2191.

Zeggini E, Scott L, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G and others. 2008. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 40:638–645.