HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2018 June 27.

Published in final edited form as:

Biometrics. 2018 June ; 74(2): 653–662. doi:10.1111/biom.12813.

Testing for Gene-Environment Interaction Under Exposure Misspecification

Ryan Sun[1,*], **Raymond J. Carroll**[2], **David C. Christiani**[3], and **Xihong Lin**[1,**]

[1]Department of Biostatistics, Harvard School of Public Health, Boston, MA, U.S.A

[2]Department of Statistics, Texas A&M University, College Station, TX, U.S.A. [2]School of Mathematical and Physical Sciences, University of Technology Sydney, Sydney, Australia

[3]Department of Environmental Health, Harvard School of Public Health, Boston, MA, U.S.A

Summary

Complex interplay between genetic and environmental factors characterizes the etiology of many diseases. Modeling gene-environment (GxE) interactions is often challenged by the unknown functional form of the environment term in the true data-generating mechanism. We study the impact of misspecification of the environmental exposure effect on inference for the GxE interaction term in linear and logistic regression models. We first examine the asymptotic bias of the GxE interaction regression coefficient, allowing for confounders as well as arbitrary misspecification of the exposure and confounder effects. For linear regression, we show that under gene-environment independence and some confounder-dependent conditions, when the environment effect is misspecified, the regression coefficient of the GxE interaction can be unbiased. However, inference on the GxE interaction is still often incorrect. In logistic regression, we show that the regression coefficient is generally biased if the genetic factor is associated with the outcome directly or indirectly. Further we show that the standard robust sandwich variance estimator for the GxE interaction does not perform well in practical GxE studies, and we provide an alternative testing procedure that has better finite sample properties.

Keywords

Asymptotic bias; Genome-Wide Environmental Interaction Studies (GWEIS); Heteroscedasticity; Model misspecification; Resampling methods; Sandwich variance

1. Introduction

Many human diseases possess an etiology which is characterized by complex relationships between genetic and environmental risk factors. Studying gene-environment (GxE)

*ryansun@fas.harvard.edu

**xlin@hsph.harvard.edu

8. Supplementary Materials

Web Appendices A-G, referenced in Sections 3, 4, and 5, are available with this paper at the *Biometrics* website on Wiley Online Library. The R package `GEint` implementing our analytical bias calculations and the BICS procedure is available for download through the CRAN repository.

interactions can help us understand biological mechanisms that cause these complex diseases (Thomas, 2010). There has been a dramatic increase in the number of Genome Wide Environmental Interaction Studies (GWEIS) over the past decade, yet remarkably, the number of replicable GxE interactions in the literature is only a handful (Aschard et al., 2012; Hutter et al., 2013). The lack of success in identifying GxE interactions is often attributed to study design issues, such as inadequate sample size and population heterogeneity (Thomas, 2010), but it also suggests limitations with current statistical methodology.

The standard approach to GWEIS performs single-marker analysis over a large number of Single Nucleotide Polymorphisms (SNPs) across the genome, repeatedly fitting a gene-environment interaction generalized linear model - e.g. linear regression for continuous traits and logistic regression for binary traits. In these models, the effect of the environmental exposure is often modeled parametrically. However we generally do not know the correct functional form of the environment covariate in the true data-generating mechanism (Aschard et al., 2012). Therefore the exposure effects can be misspecified, resulting in invalid model-based inference, as presented in the example of Cornelis et al. (2012). Environment misspecification may also cause the appearance of heteroscedasticity with respect to the exposure, which can similarly invalidate inference (Almli et al., 2014).

Our work is motivated by a GWEIS from the Harvard School of Public Health Superfund Research Project. One of the main goals of the Superfund program is to study how toxic metal exposures and genetic variants interact to affect neurodevelopment outcomes, such as Bayley Scales of Infant Development (BSID) scores, among infants. Data are available on approximately 500 infants in Bangladesh and 400 infants in Mexico. About 500,000 SNPs are used in an initial analysis, and a standard GxE interaction linear model is fit for each SNP. The QQ-plots of p-values generated by testing for GxE interaction show large departures from uniformity across many different exposures, multiple outcomes, both cohorts, and even in meta-analyses of the two cohorts together. See Figure 3 for an example. However, tests for the main effects of SNPs (G) while adjusting for exposure (ε) produce a very uniform distribution of p-values. These diagnostics suggest misspecification of the GxE mean model, as described and explained in detail by Voorman et al. (2011). Surprisingly, GxE inference which utilizes the Huber-White 'sandwich' variance estimator, a commonly-proposed remedy for incorrect inference in GWEIS (Voorman et al., 2011; Tchetgen and Kraft, 2011; Almli et al., 2014) often produces inflated p-values that show a larger departure from uniformity than p-values calculated from model-based standard errors. Again, see Figure 3 for an example. Here inflation means there is an excess of small p-values, while deflation refers to the opposite.

The impact of performing inference with misspecified models for the effects of covariates has been investigated by many authors, primarily in main effects models. The GWEIS setting is unique, because we are interested in testing a possibly misspecified interaction term and not a main effect. Additionally, we allow for confounders in the model, and these confounders may be arbitrarily misspecified. In contrast, past work primarily focuses on main effect terms and often assumes the term of interest is completely independent of other

covariates in the model. Relevant literature includes Gail, Weiand, and Piantadosi (1984), Lagakos (1988), and Begg and Lagakos (1992).

For interaction models, Vansteelandt et al. (2008) derived a set of multiply robust estimators for interaction effects, but these estimators require specification of a distribution for each SNP conditional on the other covariates in the model, a complicated task with hundreds of thousands of SNPs. Rosenblum and van der Laan (2009) and Tchetgen and Kraft (2011) studied misspecification constrained to the setting when G is completely independent of all other terms in the fitted model, including the outcome. In this scenario, they showed that the estimated GxE interaction effect will be asymptotically unbiased under the null, even under environment misspecification. While important, these findings are heavily constrained by the independence assumption. For example, under the infinitesimal model of genetic contribution to disease (Gibson, 2012), a large number of G terms are associated with an outcome, so a large proportion of tests in GWEIS will violate the assumption. Furthermore, adjusting for population stratification with genotype principal components is important in genetic association studies. The principal components will introduce regression covariates that are associated with G, another example of common practice that violates the above assumption. Our work considers arbitrary dependence among all covariates in the model and can hence incorporate confounders like principal components. We allow for misspecification of these confounders as well.

There are two main objectives to this paper. First, we provide conditions for valid GxE interaction inference under the null hypothesis of no interaction effect when the environmental exposure, and possibly other covariates, are misspecified in a generalized linear model. We perform asymptotic bias analysis to show that for a linear regression model, the estimated interaction coefficient is asymptotically unbiased under the null if the genetic factor is independent of the environment and additionally all other covariates in both the true and working model are independent of either the SNP or the environment or both. However, standard inference on the GxE interaction is incorrect even under these conditions due to biased model-based standard error estimates. In addition, we show that for a logistic regression model, the asymptotic estimate of the interaction coefficient will generally be biased under environmental misspecification when the genetic factor is associated directly or indirectly with the outcome, even under gene-environment independence. For both models we confirm that bias in the model-based standard error estimate can lead to inflated and deflated QQ-plots.

Secondly, we describe why the often-proposed sandwich variance estimator may not be a panacea for inference in practical GWEIS with moderate sample sizes. Specifically, we show that the sandwich estimate can be plagued by high variability under environmental misspecification. We propose an estimator that has better finite sample properties and illustrate its utility through both simulation and application to the Superfund dataset.

2. Exposure Misspecification in GxE Inference

2.1 Assumptions and Standard Approach

Suppose that the outcome Y_i is related to covariates G_i, ε_i, $\mathbf{Z}_i$, and $\mathbf{M}_i$ by the generalized linear model (McCullagh and Nelder, 1989)

$$g(\mu_i) = \beta_0 + \beta_G G_i + \beta_\varepsilon f(\varepsilon_i) + \beta_I G_i h(\varepsilon_i) + \mathbf{Z}_i^T \boldsymbol{\beta}_Z + \mathbf{M}_i^T \boldsymbol{\beta}_M, \quad (1)$$

where $\mu_i = E(Y_i | G_i, \varepsilon_i, \mathbf{Z}_i, \mathbf{M}_i)$. For binary outcomes, $g(\cdot)$ is the logistic link. For continuous outcomes, $g(\cdot)$ is the identity link and $\text{var}(Y_i | G_i, \varepsilon_i, \mathbf{Z}_i, \mathbf{M}_i) = \sigma^2$. Let G_i denote a discrete genetic marker and ε_i an environmental exposure variable. The additional covariates $\mathbf{Z}_i^T = (Z_{1i}, \ldots, Z_{pi})$ are correctly modeled, and the covariates $\mathbf{M}_i^T = (M_{1i}, \ldots, M_{qi})$ are subject to mismodeling. Take $f(\cdot)$ and $h(\cdot)$ to be possibly nonlinear functions of ε. While $f(\cdot)$ and $h(\cdot)$ are specified to perform theoretical bias analysis of a misspecified GxE model, we do not presume that these functions will be known in a practical analysis, and knowledge of their exact forms is not necessary for our proposed inference procedure. In vector notation, we have $\boldsymbol{\beta} = (\beta_0, \beta_G, \beta_\varepsilon, \beta_I, \boldsymbol{\beta}_Z^T, \boldsymbol{\beta}_M^T)^T$ and $\mathbf{X}_i = \left\{1, G_i, f(\varepsilon_i), G_i h(\varepsilon_i), \mathbf{Z}_i^T, \mathbf{M}_i^T\right\}^T$ so that $g(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta}$. In the context of GWEIS, which are hypothesis-generating procedures, we are most interested in inference about whether $\beta_I = 0$.

Suppose that the observed data consist of n independent and identically distributed random vectors $(Y_i, G_i, \varepsilon_i, \mathbf{Z}_i, \mathbf{W}_i)$ for $i = 1, \ldots, n$, where additional observed covariates $\mathbf{W}_i^T = (W_{1i}, \ldots, W_{ri})$ are a possibly misspecified version of $\mathbf{M}_i^T$. The only restriction we place on $\mathbf{W}_i$ is that $E(Y_i | G_i, \varepsilon_i, \mathbf{Z}_i, \mathbf{M}_i, \mathbf{W}_i) = E(Y_i | G_i, \varepsilon_i, \mathbf{Z}_i, \mathbf{M}_i)$, or in other words, the misspecified covariates do not add information about Y_i above that given by $(G_i, \varepsilon_i, \mathbf{Z}_i, \mathbf{M}_i)$.

The standard test for gene-environment interaction fits the misspecified model

$$g(\tilde{\mu}_i) = \alpha_0 + \alpha_G G_i + \alpha_\varepsilon \varepsilon_i + \alpha_I G_i \varepsilon_i + \mathbf{Z}_i^T \boldsymbol{\alpha}_Z + \mathbf{W}_i^T \boldsymbol{\alpha}_W \quad (2)$$

and performs inference on $H_0 : \alpha_I = 0$. We use $\tilde{\mu}_i$ to denote that this is a misspecified model and not the true conditional mean.

Let $\boldsymbol{\alpha} = \left(\alpha_0, \alpha_G, \alpha_\varepsilon, \alpha_I, \boldsymbol{\alpha}_Z^T, \boldsymbol{\alpha}_W^T\right)^T$ and $\tilde{\mathbf{X}}_i = \left(1, G_i, \varepsilon_i, G_i \varepsilon_i, \mathbf{Z}_i^T, \mathbf{W}_i^T\right)^T$ so that $g(\tilde{\mu}_i) = \tilde{\mathbf{X}}_i^T \boldsymbol{\alpha}$. We denote $\boldsymbol{\alpha}^*$ to be the large sample limiting value of the parameter in the fitted model and let $\hat{\boldsymbol{\alpha}}$ represent the data estimate of $\boldsymbol{\alpha}^*$.

2.2 Misspecification of the Exposure Effect May Appear as Heteroscedasticity

Almli et al. (2014) studied GxE interaction models in a post-traumatic stress disorder dataset and reported that the presence of heteroscedasticity was invalidating their inference. The

authors found that the residual variance was a function of the environment, which led a QQ-plot to show heavily inflated p-values when performing genome-wide interaction testing with the standard GxE model. We show in this section that for linear regression, misspecification of the exposure effect may cause the appearance of heteroscedasticity in the environment as reported by Almli et al. (2014).

Suppose the true linear GxE interaction model is given by

$$Y_i = \beta_0 + \beta_G G_i + \beta_\varepsilon f(\varepsilon_i) + \beta_I G_i h(\varepsilon_i) + \epsilon_i,$$

where $f(\cdot)$ and $h(\cdot)$ are non-linear functions of ε_i, and $\epsilon_i \sim N(0, \sigma^2)$. Assume we fit the misspecified GxE interaction model with a linear effect of ε_i:

$$Y_i = \alpha_0 + \alpha_G G_i + \alpha_\varepsilon \varepsilon_i + \alpha_I G_i \varepsilon_i + e_i. \quad (3)$$

One can easily show that under the misspecified model (3),

$$E(e_i^2 \mid G_i, \varepsilon_i) = \sigma^2 + d(G_i, \varepsilon_i),$$

where $d(G_i, \varepsilon_i) = \{\mu_i(G_i, \varepsilon_i) - \mu_{i,mis}(G_i, \varepsilon_i)\}^2$, and

$$\mu_i(G_i, \varepsilon_i) = \beta_0 + \beta_G G_i + \beta_\varepsilon f(\varepsilon_i) + \beta_I G_i h(\varepsilon_i)$$
$$\mu_{i,mis}(G_i, \varepsilon_i) = \alpha_0 + \alpha_G G_i + \alpha_\varepsilon \varepsilon_i + \alpha_I G_i \varepsilon_i.$$

If $f(\cdot)$ is not linear in ε_i, then the function $d(G_i, \varepsilon_i)$ is generally not 0 even under the null hypothesis $\beta_I = 0$. Thus there will appear to be heteroscedasticity with respect to the effect of the environment. This example suggests that it is possible GxE studies of continuous outcomes may misdiagnose exposure misspecification as heteroscedasticity; such studies may also find the following results relevant to their work.

3. Inference in the Misspecified Model

3.1 Asymptotic Bias of Fitted Coefficients for the Identity Link

Although our primary concern lies in testing α_I, it is often also of interest to estimate other parameters in $\boldsymbol{\alpha}$ for interpretability reasons or joint tests such as the 2-df test of $H_0 : \alpha_G = \alpha_I = 0$ proposed by Kraft et al. (2007) and utilized by Almli et al. (2014).

When $g(\cdot)$ is the identity link, the $p + r + 4$ score equations for estimating $\boldsymbol{\alpha}$ under the fitted model (2) are

$$(n\sigma^2)^{-1}\sum_{i=1}^{n}(1, G_i, \varepsilon_i, G_i\varepsilon_i, \mathbf{Z}_i^T, \mathbf{W}_i^T)^T\left(Y_i - \tilde{\mathbf{X}}_i^T\boldsymbol{\alpha}\right) = \mathbf{0}_{(p+r+4)\times 1}.$$

The asymptotic limit $\boldsymbol{\alpha}$ of the MLE $\hat{\boldsymbol{\alpha}}$ is the value such that

$$E\left\{(1, G, \varepsilon, G\varepsilon, \mathbf{Z}^T, \mathbf{W}^T)^T\left(\mathbf{X}^T\boldsymbol{\beta} - \tilde{\mathbf{X}}^T\boldsymbol{\alpha}\right)\right\} = \mathbf{0}_{(p+r+4)\times 1}. \quad (4)$$

Under distributional assumptions, it is possible to solve equation (4) in closed form and find the asymptotic bias of each fitted covariate. In derivations for this section, we will assume without loss of generality that the covariates G and ε are centered at 0. Also, subscripts on μ will denote the expectation of those subscripts, so that $\mu_{G\varepsilon} = E(G\varepsilon) = \text{Cov}(G, \varepsilon)$.

The asymptotic value of α_I takes the general form

$$\alpha_I = \beta_\varepsilon * C_1 + \beta_I * C_2 + \mathbf{C}_3^T\boldsymbol{\beta}_M,$$

where (C_1, C_2) denote constants and $\mathbf{C}_3$ denotes a $q\times 1$ vector of constants. These constants depend on the form of the misspecification as well as the marginal and joint distribution of the covariates. Similarly, α_G is also a complicated function of the true effect sizes. Under the null, we can perform valid inference on α_I if $C_1 = 0$ and $\mathbf{C}_3 = \mathbf{0}_{q\times 1}$. The same is true for the constants relating to α_G. The full expansions are unwieldy and difficult to examine, so we leave them to Web Appendix A. Web Appendix B also explores how these equations can be extended to consider joint testing for interactions between the exposure and a set of SNPs, e.g., SNPs in a gene.

In the following paragraphs, we briefly highlight some of the most interpretable consequences of the equations and describe the implications on GWEIS study design. For an arbitrary set of covariates and dependence structures, we offer an R package `GEint` that is able to calculate the exact magnitude of bias in fitted coefficients, given some inputs on the true model. This software offers a very flexible platform for users to analyze bias on a case-by-case basis, and it can also be used, for example, to perform sensitivity analysis on GWEIS models.

- Consider first a simple testing case where only the environment term is misspecified in the fitted model, that is, $\mathbf{W} = \mathbf{M} = 0$. Under $H_0 : \beta_I = 0$, sufficient conditions for $\alpha_I = 0$ are gene-environment independence combined with

$$\mu_{G\varepsilon Z_1} = \ldots = \mu_{G\varepsilon Z_p} = 0. \quad (5)$$

The sufficient conditions are achieved under gene-environment independence and if at least one of G or ε is independent of each Z_j for all $j = 1, ..., p$.

- Additionally, under the joint null $H_0 : \beta_G = \beta_I = 0$, sufficient conditions for $\alpha_G = \alpha_I = 0$ are gene-environment independence combined with

$$\mu_{GZ_1} = \ldots = \mu_{GZ_p} = 0. \quad (6)$$

The sufficient condition (6) is achieved if G is independent of each Z_j for all $j = 1, ...p$, which is much more stringent. This result suggests the joint test is much more susceptible to issues of bias due to model misspecification.

- Next consider the case where other covariates are also misspecified, so that $\boldsymbol{W} \neq \boldsymbol{M}$. Under the null $H_0 : \beta_I = 0$, two sufficient conditions for $\alpha_I = 0$ are gene-environment independence combined with

$$\mu_{G\varepsilon Z_1} = \ldots = \mu_{G\varepsilon Z_p} = \mu_{G\varepsilon M_1} = \ldots = \mu_{G\varepsilon M_q} = \mu_{G\varepsilon W_1} = \ldots = \mu_{G\varepsilon W_r} = 0. \quad (7)$$

The sufficient condition (7) is achieved if at least one of ε or G is independent of each $(Z_1...Z_p)$, each $(M_1...M_q)$, and each $(W_1...W_r)$ in addition to gene-environment independence. The result of Rosenblum and van der Laan (2009) is a special case of this result.

- Under the joint null hypothesis $H_0 : \beta_G = \beta_I = 0$, two sufficient conditions for $\alpha_G = \alpha_I = 0$ are gene-environment independence combined with

$$\mu_{GZ_1} = \ldots = \mu_{GZ_p} = \mu_{GM_1} = \ldots = \mu_{GM_q} = \mu_{GW_1} = \ldots = \mu_{GW_r} = 0. \quad (8)$$

The sufficient condition (8) is achieved if G is independent of each $(Z_1, ...,Z_p)$, each $(M_1, ...,M_q)$, and each $(W_1, ..., W_r)$.

The scenarios discussed above suggest that when genetic and environmental covariates are independent, GWEIS inference is likely to be more robust to model misspecification. When G and ε are dependent and the effect of ε is misspecified, the estimate of the interaction term will often be asymptotically biased. In addition, the results suggest that introducing many additional covariates into the model, for instance to reduce the standard error of estimated coefficients, is likely to increase the chance of model misspecification and cause biased inference on GxE interactions.

3.2 Controlling for Population Stratification

Population stratification due to heterogeneous populations is common in genetic association studies and is routinely adjusted for by introducing genetic principal components as covariates. In the presence of population stratification and use of principal components, the results in (7) suggest that if the environmental exposure varies with sub-populations,

misspecification of the exposure is likely to result in biased inference on the GxE interaction. For instance, if a cohort is composed of Northern and Southern Europeans, and the exposure of interest is differentiated between these two sub-population groups, neither the environmental term nor the genetic term will be independent of the principal components covariates. Then inference on α_I is likely to be sensitive to misspecification of the exposure effects.

3.3 Asymptotic Bias of Fitted Coefficients for the Logistic Link

For binary outcomes and a logistic regression model, the score equations become:

$$0 = n^{-1} \sum_{i=1}^{n} (1, G_i, \varepsilon_i, G_i\varepsilon_i, \boldsymbol{Z}_i^T, \boldsymbol{W}_i^T)^T \left\{ Y_i - \mu_i(\widetilde{\mathbf{X}}_i^T \boldsymbol{\alpha}) \right\}, \quad (9)$$

where $\mu(x) = g^{-1}(x) = \exp(x)/\{1 + \exp(x)\}$. The asymptotic limit $\boldsymbol{\alpha}$ is the value such that

$$E\left[(1, G, \varepsilon, G\varepsilon, \boldsymbol{Z}^T, \boldsymbol{W}^T)^T \left\{ \mathbf{Y} - \mu(\widetilde{\mathbf{X}}\boldsymbol{\alpha}) \right\} \right] = \mathbf{0}. \quad (10)$$

These equations generally do not have closed forms. Rosenblum and van der Laan (2009) and Tchetgen and Kraft (2011) studied the case where $\beta_G = \beta_I = 0$ and G is independent of all other terms in the true model (1). The authors showed that in this setting, $\alpha_G = \alpha_I = 0$ even under environment misspecification.

Here we focus on situations where the independence assumptions do not hold. We perform asymptotic bias calculations by numerically solving (10) for specific cases to demonstrate that when the Rosenblum and van der Laan conditions are not met, α_I will likely be biased. That is, if G has some association with Y and the effect of the environment is misspecified, then α_I is generally biased away from 0 under the null.

Figure 1 illustrates the asymptotic bias in the interaction term under four different misspecification scenarios. To be completely clear, we assume in these scenarios and for the rest of this section that the main effect of the exposure exists in the true model and has been misspecified in the fitted model. In all cases, we solve the asymptotic score equations (10) using numerical methods. For each setting we assume that G has a Binomial(2, 0.3) distribution and that it is correlated with the underlined variable (see Figure 1 for definition) by an amount given on the x-axis. The variable correlated with G is assumed to be a mixture of normal random variables, with mean conditional on G, and it has marginal mean 0 and variance 1. In scenarios 2–4 of Figure 1, the environment term is independently generated as a standard normal random variable. In scenario 3, $M = W^2$ provides additional misspecification.

We see that in scenario 1, the interaction coefficient is biased because G is associated with ε, which is in the true model. Thus when there is no gene-environment independence, the interaction coefficient will be biased. In scenarios 2 and 3, the interaction coefficient is

biased because G is indirectly associated with Y through correlation with Z and W respectively. Thus if the true model includes principal components to control for population stratification, α_I will be biased. In scenario 4, G is correlated with W, but W has no association with terms in the true model, so there is no bias.

These four scenarios cover a wide range of possibilities, and they show that the estimate of the interaction term is generally biased under environment misspecification if the genetic term is directly or indirectly associated with the outcome. In Web Appendix C we are able to provide some more intuition on how bias arises in the simplest situations where there are no additional covariates. If G is not directly or indirectly associated with the outcome Y through correlation with other terms in the true model, then the interaction regression coefficient will be asymptotically unbiased under the null.

3.4 Asymptotic Standard Error of Fitted Coefficients

Most earlier work on model misspecification (Voorman et al., 2011; Tchetgen and Kraft, 2011) advocates that using a robust sandwich standard error estimate will provide asymptotically correct Type I error when α_I is unbiased under the null. The same theory holds for the models we study, because the asymptotic covariance matrix of $\hat{\boldsymbol{\alpha}}$ is given by:

$$V_{\hat{\boldsymbol{\alpha}}} = B(\boldsymbol{\alpha})^{-1} A(\boldsymbol{\alpha}) \left\{ B(\boldsymbol{\alpha})^{-1} \right\}^T;$$
$$B(\boldsymbol{\alpha}) = E\left\{ \frac{\partial \boldsymbol{\psi}(\tilde{\boldsymbol{X}}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^T} \right\}, A(\boldsymbol{\alpha}) = E\left\{ \boldsymbol{\psi}(\tilde{\boldsymbol{X}}, \boldsymbol{\alpha}) \boldsymbol{\psi}(\tilde{\boldsymbol{X}}, \boldsymbol{\alpha})^T \right\},$$

where $\boldsymbol{\psi}(\tilde{\boldsymbol{X}}, \boldsymbol{\alpha})$ are the $p + r + 4$ score equations from above. The model-based variance estimator assumes that $B(\boldsymbol{\alpha}) = -A(\boldsymbol{\alpha})$, which is incorrect under exposure misspecification and will invalidate the inference, even if the regression coefficient estimate is unbiased.

Denote by $\hat{m}_{\alpha_I}$ the model-based standard error estimate of $\hat{\alpha}_I$. We show in Web Appendix D that the model-based Wald statistic for testing the interaction term $T_{mod} = (\hat{\alpha}_I / \hat{m}_{\alpha_I})^2$ asymptotically follows a scaled chi-square distribution $c\chi^2_1$, where the expressions of c for linear and logistic regression are given in that appendix. If $c > 1$ for many SNPs across the genome, then the QQ-plot for GxE interactions using model-based standard errors will show inflated p-values. If $c < 1$ for many SNPs, then the QQ-plot will show deflated p-values. The value of c converges to a figure which is determined by both the true and fitted models.

Using a sandwich estimator with $\hat{\boldsymbol{\alpha}}$ instead of $\boldsymbol{\alpha}$ in the above expression provides a consistent variance estimate. However, as noted above, when we utilized this strategy on our Superfund dataset, p-values calculated with the robust standard error sometimes seemed less uniform than p-values calculated with the model-based standard error.

4. Alternative Standard Error Estimates

4.1 Inflation Caused by the Sandwich Estimator

Even though many studies suggest to use the robust sandwich variance estimator, the Superfund data ($n \approx 400$), the study of Almli et al. (2014) ($n > 3000$), and the analysis of

Cornelis et al. (2012) ($n > 3000$) are a few examples where inference conducted with the sandwich estimator appears to return an excess of highly significant p-values. It is known that the sandwich estimator is often biased downwards and is more variable than model-based estimators (Kauermann and Carroll, 2001) even when the model is not misspecified, which can cause inflated Type I error in hypothesis testing (Kauermann and Carroll, 2001).

Exposure misspecification can exacerbate the variability of the sandwich estimator in linear regression. This occurs because the sandwich estimator is a linear combination of the squared regression residuals, and the squared regression residuals have more variance under exposure misspecification. We demonstrate in detail in Web Appendix E how the variance of the sandwich estimator can be much larger under model misspecification than when the model is correctly specified. The natural downward bias of the sandwich estimator as well the additional variability caused by exposure misspecification provide intuition for the heavily inflated sandwich p-values seen in the Superfund data.

A similar derivation incorporating residual variability in logistic regression is complicated by the difficulty of specifying a distribution for the residuals. However, in our simulations, we find that testing for binary outcomes with the sandwich standard error can have slightly incorrect size as well. Thus it is of interest to find variance estimators which can better protect the level of the test when performing inference under exposure misspecification.

4.2 Bootstrap Inference with a Corrected Sandwich

As an alternative to the model-based and sandwich variance estimators, we propose a resampling-based method. The proposed method can be thought of as a finite sample correction to the sandwich estimator. Denote by $T_{sand} = (\hat{\alpha}_I / \hat{s}_{\hat{\alpha}_I})^2$ a test statistic for the interaction effect calculated using the sandwich standard error estimate $\hat{s}_{\hat{\alpha}_I}$. This test statistic should asymptotically have a χ^2_1 distribution under the null. If the sandwich estimator is biased in finite samples, then the bias will cause the test statistic to instead have an approximately scaled chi-square distribution: $T_{sand} \approx c\chi^2_1$. We can approximate the $c\chi^2_1$ distribution by resampling the test statistic and matching the moments of its sampling distribution with a Satterthwaite-type idea as follows:

Fit model (2) on the observed data to find the estimated interaction coefficient $\hat{\alpha}_I^{(init)}$ and sandwich test statistic $T_{sand}^{(init)}$. For each of $b = 1, 2, ..., B$, say $B = 1000$, bootstrap iterations, perform a nonparametric bootstrap by sampling $(Y_i, \tilde{\boldsymbol{X}}_i)$ from the original data n times with replacement. Fit model (2) on the new sample. Calculate the squared, centered bootstrap test statistics $T_{sand}^{(b)} = \{(\hat{\alpha}_I^{(b)} - \hat{\alpha}_I^{(init)})/\hat{s}_{\hat{\alpha}_I}^{(b)}\}^2$ where $\hat{\alpha}_I^{(b)}$ and $\hat{s}_{\hat{\alpha}_I}^{(b)}$ are the regression coefficient and the sandwich standard error estimate for the interaction term based on the bth bootstrap sample. Match the mean and variance of $\mathbf{T} = \left(T_{sand}^{(1)}, \ldots, T_{sand}^{(B)}\right)$ to the moments of a $k\chi^2_a$ distribution, where we solve for (k, a) using the equations $k = \text{Var}(\mathbf{T})/\{2 * \text{Mean}(\mathbf{T})\}$ and $a = \text{Mean}(\mathbf{T})/k$. Find the p-value of the original test statistic $T_{sand}^{(init)}$ using $k\chi^2_a$ as the reference distribution. We will refer to this method as the Bootstrap Inference with Corrected

Sandwich (BICS) procedure; an extension to the joint testing case with multiple interaction terms is discussed in Web Appendix F. We also note that a natural alternative, using the empirical standard error of $\hat{\alpha}_I^{(1)}, \ldots, \hat{\alpha}_I^{(B)}$ instead of the sandwich estimate, does not work well.

5. Simulation Studies

We conducted a variety of simulations to evaluate control of Type I error rate in GWEIS for different testing procedures over a range of misspecification scenarios. All misspecified models we consider are generated under the null of $\beta_I = 0$, and all satisfy the conditions for valid inference discussed previously, that is, $\alpha_I = 0$ asymptotically. The Type I error rate of the tests should be controlled at the nominal size of 0.05 with an unbiased standard error estimator. In all simulations we fit model (2) with $\boldsymbol{Z}_i = \boldsymbol{W}_i = 0$. We use a Wald t-test to generate p-values with the naive and sandwich standard errors. Each misspecified model is tested at sample sizes from 400 to 3200 to reflect the finite sample problem which affects the Superfund study. We perform 50,000 replications of the simulation at each parameter setting and report the percentage of times that each testing procedure rejects the null.

We first describe the misspecification for continuous outcomes. Simulation A has outcome Y generated from the model $Y_i = \beta_\varepsilon \varepsilon_i^3 + \epsilon_i$, $\epsilon_i \sim N(0,1)$ where β_ε is chosen such that ε explains 10% of the variance in Y. In Simulation B we increase the degree of misspecification by taking the true model to be $Y_i = \beta_\varepsilon \varepsilon_i^3 + \epsilon_i$, $\epsilon_i \sim N(0,1)$, where β_ε is again chosen such that ε explains 10% of the variance in Y. For both Simulations A and B, we generate $\varepsilon_i \sim N(1, 1)$. Simulations C and D have the same true model as A and B, except we generate $\varepsilon_i \sim \text{Beta}(2,5)$ to introduce skewness into the exposure variable. We also adjust β_ε so that ε continues to explain 10% of the variance in Y. Finally, Simulation E differs from the previous four in that we generate the outcome as $Y_i = \beta_G G_i + \beta_\varepsilon \varepsilon_i^2 + \epsilon_i$, with ε_i and ϵ_i again as they were in Simulation A. This situation mimics testing for interaction with a SNP that has a marginal effect but no interaction effect. The values of β_G and β_ε are chosen such that ε and G would explain 10% and 1% of the variance in Y respectively if G had minor allele frequency 0.3. For all scenarios above, G is simulated by using HAPGEN2 to generate the number of minor alleles at a random SNP on chromosome 1 (HapMap3 CEU population used as reference), thus G and ε are always independent.

We see from Table 1 that the sandwich estimator often produces inflated Type I error rates. BICS performs very well, protecting the size almost exactly in every single situation. Of course, the sandwich estimator performs progressively better as the sample size increases. In contrast, BICS does not appear to show a trend in n and increases its relative superiority over the sandwich estimator at the smallest sample sizes. The naive estimator is always biased and shows the most inflation. These results closely reflect the trends in our data example, where QQ-plots of p-values calculated with the sandwich and naive estimators show very early departures from the 45-degree line, indicating lack of uniformity.

Next we consider binary outcomes. Simulations F,G,H,I, and J are conducted in the same spirit as the previous five. The outcome Y_i in simulation F is generated from the model:

$$Y_i \sim \text{Bernoulli}(\pi_i);\ \pi_i = \frac{\exp(0.4\varepsilon_i^2)}{1 + \exp(0.4\varepsilon_i^2)}.$$

The parameter $\beta_0 = 0$ is chosen to give a subject with $\varepsilon = 0$ a disease probability of 0.5. Simulation G is conducted under a higher degree of misspecification as we take the true probability of disease to be $\pi_i = \exp(0.2\varepsilon_i^3)/\{1 + \exp(0.2\varepsilon_i^3)\}$. For simulations F and G we generate $\varepsilon_i \sim N(0, 1)$. Simulations H and I have $\pi_i = \exp(\varepsilon_i^2)/\{1 + \exp(\varepsilon_i^2)\}$ and $\pi_i = \exp(\varepsilon_i^3)/\{1 + \exp(\varepsilon_i^3)\}$ respectively with $\varepsilon_i \sim$ Beta(2,5). Finally, in Simulation J each SNP has a marginal effect with $\pi_i = \exp(0.1G_i + 0.4\varepsilon_i^2)/\{1 + \exp(0.1G_i + 0.4\varepsilon_i^2)\}$, and $\varepsilon_i \sim N(0, 1)$ again.

In these logistic regression simulations we see that the sandwich estimator actually performs fairly well, with the correct size in most situations. It can be slightly conservative when $n = 400$. BICS similarly performs well, although it appears to be slightly less conservative than using the sandwich estimate. In absolute terms, BICS and the sandwich estimator both appear to deviate a similar amount from the expected size. Once again the naive standard error estimate is biased and produces tests at an incorrect size.

Our simulation study suggests that BICS protects against inflated p-values in linear regression GWEIS of moderate size. In the simulations and in published articles, inflation from inference with the sandwich estimator appears to occur even at $n = 3000$; while each study is unique, we would use this figure as a rough benchmark for moderate size. Although our simulations consider a finite number of settings, our results show that the trends presented above appear to hold for a variety of different scenarios. When n is large or logistic regression is used, we agree with previous suggestions that the sandwich estimator should be employed for its speed and simplicity, however BICS can be used as an alternative if diagnostic QQ-plots appear worrisome. A simple and fast implementation is available through `GEint`. Additional power simulations are available in Web Appendix G.

We also note that in practical setting, the environment term will remain constant for each SNP, while in our simulation the environment term is newly generated with each different SNP. This choice was made to present the fairest possible comparison in simulation. When the environment term is held constant for each SNP, the difference between BICS and the sandwich estimator can be even more drastic (again see Figure 3).

6. Application to Superfund Data

One major goal of the Superfund Research Program is to study the interplay of genes and toxic metal exposures on childhood neurological outcomes. The metal exposure of interest is lead concentration in the umbilical cord blood. The neurological outcome is a mental composite score calculated from the BSID. There exists evidence that exposure to certain

metals during the prenatal period can seriously impair the cognitive development of infants (Claus Henn et al., 2012), but to date we are unaware of any previous gene-environment interaction studies covering toxic metal exposures and neurodevelopment outcomes.

The participants enrolled in the study come from two cohorts. Recruitment in Mexico was described in Burris et al. (2013), with 389 of the recruited mother-infant pairs having complete genetic and covariate data. Recruitment in Bangladesh was described in Kile et al. (2014), with 497 pairs having complete genetic and covariate data. Briefly, women were enrolled during hospital visits in the early weeks of their pregnancy, and covariate information was collected upon subsequent visits to the hospital. Genotyping was performed using the Illumina OmniExpressExome-8 in the Bangladesh cohort and the Illumina HumanOmni1-Quad Beadchip in the Mexico cohort. About 500,000 SNPs common to both cohorts remained after quality control.

We conducted a standard GWEIS by repeatedly fitting the model

$$Y_i = \alpha_0 + \alpha_G G_i + \alpha_\varepsilon \varepsilon_i + \alpha_I G_i \varepsilon_i + \boldsymbol{Z}_i^T \boldsymbol{\alpha}_Z + \epsilon_i, \quad (11)$$

where $\boldsymbol{Z}_i$ is an 8×1 vector of additional covariates including sex, birthweight, gestational age, education of mother (binary, 1 if primary school or greater), household environmental smoke (binary), child's age at time of assessment, and the first two genotype principal component vectors. Here ε is the logarithm of umbilical cord blood lead concentration. Two distinct genome-wide scans were conducted, one for each cohort. A meta-analysis was then performed to pool the data and provide a final measure of association for each SNP.

The initial analyses implemented with a model-based standard error produced QQ-plots of highly non-uniform p-values (Figure 2 and Figure 3). We conjectured that a major cause of the non-uniformity was misspecification of the effect of the environmental covariate. To investigate possible misspecification, we repeated the initial analyses but introduced a spline term for the environment instead of modeling it linearly. QQ-plots produced after this modification improved somewhat but still showed some non-uniformity. We also performed a standard GWAS by removing the interaction term from the fitted model and only testing for the marginal effect of G. QQ-plots for the GWAS seemed relatively uniform.

Under model misspecification, the theoretical results derived in Section 3 suggest that we can have robust tests of the null hypothesis under some independence conditions, which we believe are reasonable to assume here. Still, as shown in Figures 2 and 3, the QQ-plots based on sandwich standard errors are inflated.

We next re-analyze the data by fitting model (11) and using BICS to generate p-values for the cohort-specific GWEIS. For the meta-analysis, we use METAL (Willer et al., 2010) to combine the BICS p-values from each cohort. After meta-analysis, the resampling-based p-values are much more uniform than p-values calculated using the naive or sandwich variance estimate. It appears that our assumptions about independence mostly hold, as there is little inflation using BICS. The corrected p-values seem to reflect that $\alpha_I = 0$ throughout much of

the genome, and inflation seen from using the sandwich estimate can likely be attributed to the drawbacks discussed in Section 4.

After applying BICS and accounting for multiple testing, we do not find any SNPs to be significant at the genome-wide level in either of the cohorts. No SNPs reach genome-wide significance in the meta-analysis either. However, the meta-analysis does suggest a promising region for future study. Two of the top SNPs identified in the meta-analysis are rs9642758 and rs10503970 (p-values of 8.79×10^{-6} and 2.57×10^{-5} respectively), which are both located on chromosome 8 in the region of the gene UNC5D. UNC5D encodes a receptor for netrin, which may be involved in axon guidance and could plausibly affect infant neurodevelopment through interaction with toxic metals. We believe the interaction between UNC5D, exposure to lead, and neurodevelopment outcomes is a promising candidate for further study.

7. Discussion

It is often the case in the standard GWEIS approach that the parametric form of the environmental covariate will be misspecified. We have demonstrated conditions under which inference for the interaction effect is still valid under model misspecification. These results provide guidance on fitting GxE interaction models. We show that for linear regression models, the estimate of the interaction effect will be asymptotically unbiased if there is both gene-environment independence and also either the genetic or environment term is independent of each coefficient in both the true and fitted models. For logistic regression models, the estimate of the interaction effect will generally only be asymptotically unbiased if the genetic term is neither directly nor indirectly associated with the outcome.

When the conditions for valid inference on GxE interactions are met, hypothesis testing may still be difficult to conduct because the model-based estimate of standard error is biased under environment misspecification, and the Huber-White sandwich estimator can lead to excess Type I error in finite samples. We provide a resampling-based method of obtaining p-values and show its advantages both in simulation and through application to the Superfund dataset; BICS provides an especially useful inference tool in linear regression GWEIS with moderate sample sizes. After reanalysis of the Superfund data, we have identified UNC5D as a strong candidate gene for further study in how lead exposure can affect infant neurodevelopment.

While our resampling method can be computed rather quickly and has been found to work well in practical studies of moderate sample size, it is still a minor drawback to perform a bootstrap procedure for every SNP across the genome. It is of future research interest to develop more computationally efficient inference methods for robust testing of GxE interactions in GWEIS. Similarly, it is desirable to develop semiparametric gene-environment interaction models that are more robust to model misspecification in both the exposure and confounder covariates. For example, Maity et al. (2009) introduced a semiparametric interaction model with a nonparametrically-modeled exposure.

We additionally showed in this paper that our asymptotic bias analysis and resampling inference method are applicable to SNP-set by environment interaction models when the number of SNPs in a set is not large. When the number of SNPs in a set is large, for example when considering SNPs in a genetic pathway, Lin et al. (2013) proposed a variance component test for SNP-set by environment interaction that demonstrated attractive performance in joint testing of many interaction terms. It is also of future research interest to study the validity of such variance component interaction tests when exposure effects are misspecified.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the associate editor and two referees for their valuable comments. This work is supported by NIH grants R35-CA197449, P01-CA134294, R01-HL113338, P42-ES016454, U01-CA057030, and T32-ES007142.

References

Almli L, Duncan R, Feng H, Ghosh D, Binder E, Bradley B, Ressler K, Conneely K, Epstein M. Correcting systematic inflation in genetic association tests that consider interaction effects: application to a genome-wide association study of posttraumatic stress disorder. JAMA Psychiatry. 2014; 71:1392–1399. [PubMed: 25354142]

Aschard H, Lutz S, Maus B, Duell E, Fingerlin T, Chatterjee N, Kraft P, Van Steen K. Challenges and opportunities in genome-wide environmental interaction (GWEIS) studies. Human Genetics. 2012; 131:1591–1613. [PubMed: 22760307]

Begg M, Lagakos S. Effects of mismodeling on tests of association based on logistic regression models. Annals of Statistics. 1992; 20:1929–1952.

Burris H, Braun J, Byun H, Tarantini L, Mercado A, Wright R, Schnaas L, Baccarelli A, Wright R, Tellez-Rojo M. Association between birth weight and dna methylation of IGF2, glucocorticoid receptor and repetitive elements line-1 and alu. Epigenomics. 2013; 5:271–281. [PubMed: 23750643]

Claus Henn B, Schnass L, Ettinger A, Schwartz J, Lamadrid-Figueroa H, Hernndez-Avila M, Amarasiriwardena C, Hu H, Bellinger D, RW, Tellez-Rojo M. Associations of early childhood manganese and lead coexposure with neurodevelopment. Environmental Health Perspectives. 2012; 120:126–131. [PubMed: 21885384]

Cornelis M, Tchetgen E, Liang L, Qi L, Chatterjee N, Hu F, Kraft P. Gene-environment interactions in genome wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. American Journal of Epidemiology. 2012; 120:191–202.

Gail M, Weiand S, Piantadosi S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika. 1984; 71:431–444.

Gibson G. Rare and common variants: twenty arguments. Nature Reviews Genetics. 2012; 13:135–145.

Hutter C, Mechanic L, Chatterjee N, Kraft P, Gillanders E. Gene-environment interactions in cancer epidemiology: a National Cancer Institute think tank report. Genetic Epidemiology. 2013; 37:643–657. [PubMed: 24123198]

Kauermann G, Carroll R. A note on the efficiency of sandwich covariance matrix estimation. Journal of the American Statistical Association. 2001; 96:1387–1396.

Kile M, Rodrigues E, Mazumdar M, Dobson C, Diao N, Golam M, Quamruzzaman Q, Rahman M, Christiani D. A prospective cohort study of the association between drinking water arsenic

exposure and self-reported maternal health symptoms during pregnancy in bangladesh. Environmental Health. 2014; 13:29. [PubMed: 24735908]

Kraft P, Yen Y, Stram D, Morrison J, Gauderman W. Exploiting gene environment interaction to detect genetic associations. Human Heredity. 2007; 63:111–119. [PubMed: 17283440]

Lagakos S. Effects of mismodelling and mismeasuring explanatory variables on tests of their association with a response variable. Statistics in Medicine. 1988; 7:257–274. [PubMed: 3353607]

Lin X, Lee S, Christiani D, Lin X. Tests for interactions between a genetic marker set and environment in generalized linear models. Biostatistics. 2013; 14:667–681. [PubMed: 23462021]

Maity A, Carroll R, Mammen E, Chatterjee N. Testing in semiparametric models with interaction, with applications to gene-environment interactions. Journal of the Royal Statistical Society: Series B. 2009; 71:75–96.

McCullagh, P., Nelder, JA. Generalized Linear Models. CRC press; 1989.

Rosenblum M, van der Laan M. Using regression models to analyze randomized trials: asymptotically valid hypothesis tests despite incorrectly specified models. Biometrics. 2009; 65:937–945. [PubMed: 19210739]

Tchetgen E, Kraft P. On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. Epidemiology. 2011; 22:257–261. [PubMed: 21228699]

Thomas D. Gene-environment-wide association studies: emerging approaches. Nature Reviews Genetics. 2010; 11:259–272.

Vansteelandt S, Vander Weele T, Tchetgen E, Robins J. Multiply robust inference for statistical interactions. Journal of the American Statistical Association. 2008; 103:1693–1704. [PubMed: 21603124]

Voorman A, Lumley T, McKnight B, Rice K. Behavior of qq-plots and genomic control in studies of gene-environment interaction. PLOS ONE. 2011; 6:e19416. [PubMed: 21589913]

Willer CJ, Li Y, Abecasis G. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]

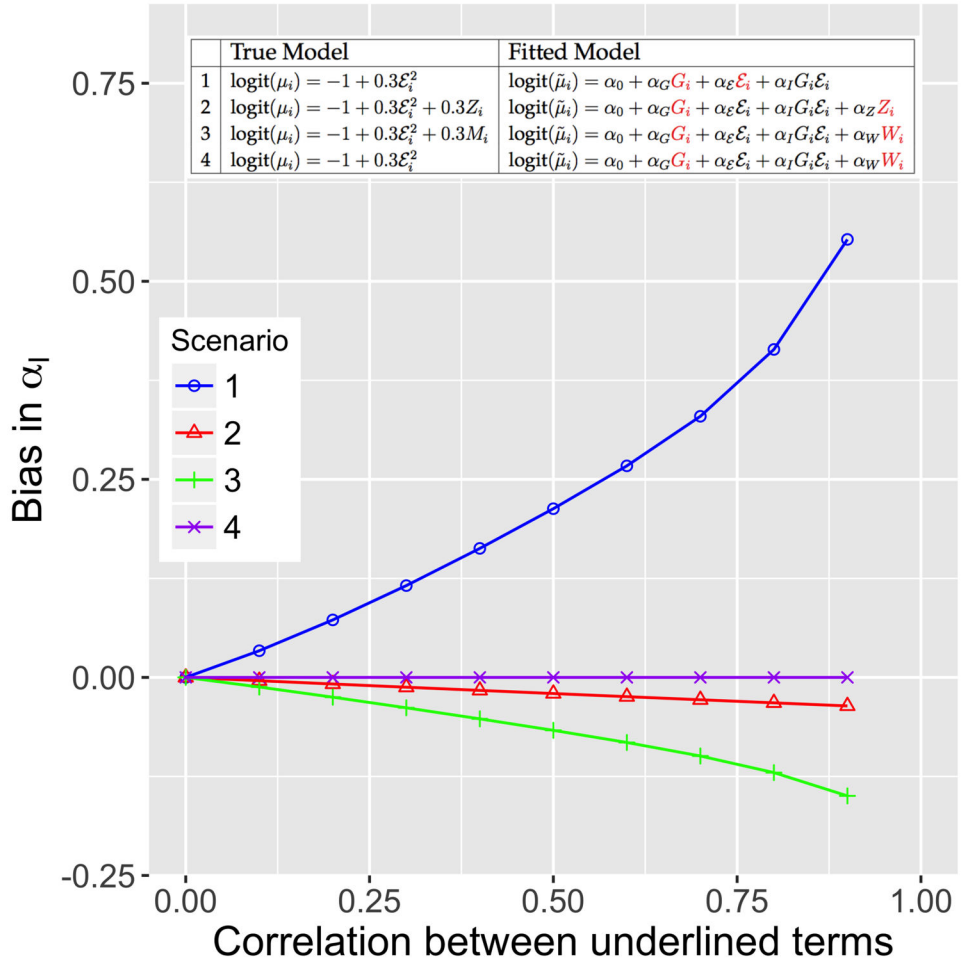

Figure 1.

Bias of the fitted interaction coefficient in logistic regression over four different misspecification settings. The underlined terms are correlated, with the magnitude of correlation given on the x-axis. At each data point we solved the score equations numerically and confirmed these results through simulation.

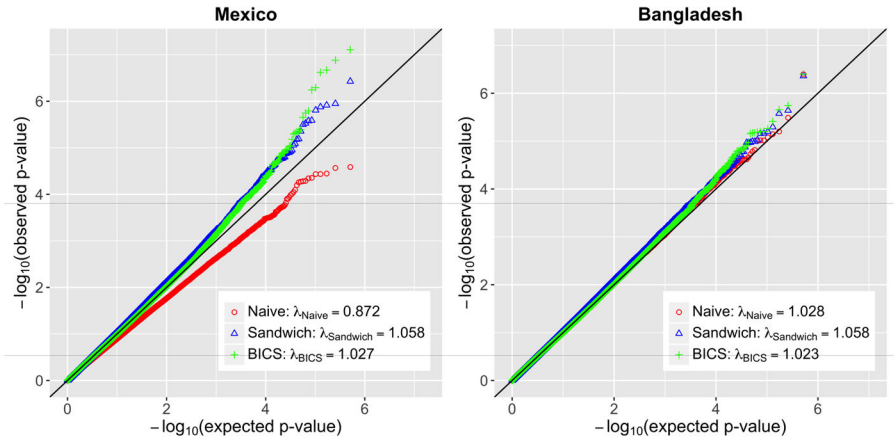

Figure 2.

QQ-plots of p-values generated by testing for interaction effect in model (11) with naive model-based variance, sandwich variance, and BICS procedure. The outcome is mental composite score calculated from the BSID. The exposure is logarithm of umbilical cord blood lead concentration. We expect to see close adherence to the 45-degree line through the left half of the x-axis. On the right half of the x-axis are less than 0.1% of all terms; these show the most evidence of association and may indicate true signals. However in both cohorts the sandwich and naive p-values show very early departures from the 45-degree line. Such behavior is worrisome because it indicates the inference procedure is not producing uniform p-values under the null, and thus all inferences we make may be invalidated. A quantitative measure of the departure from uniformity is given by the genomic inflation factor, provided in the legend. This factor is defined as the ratio of the median of the empirically observed test statistics to the expected median of a chi-squared distribution with one degree of freedom.

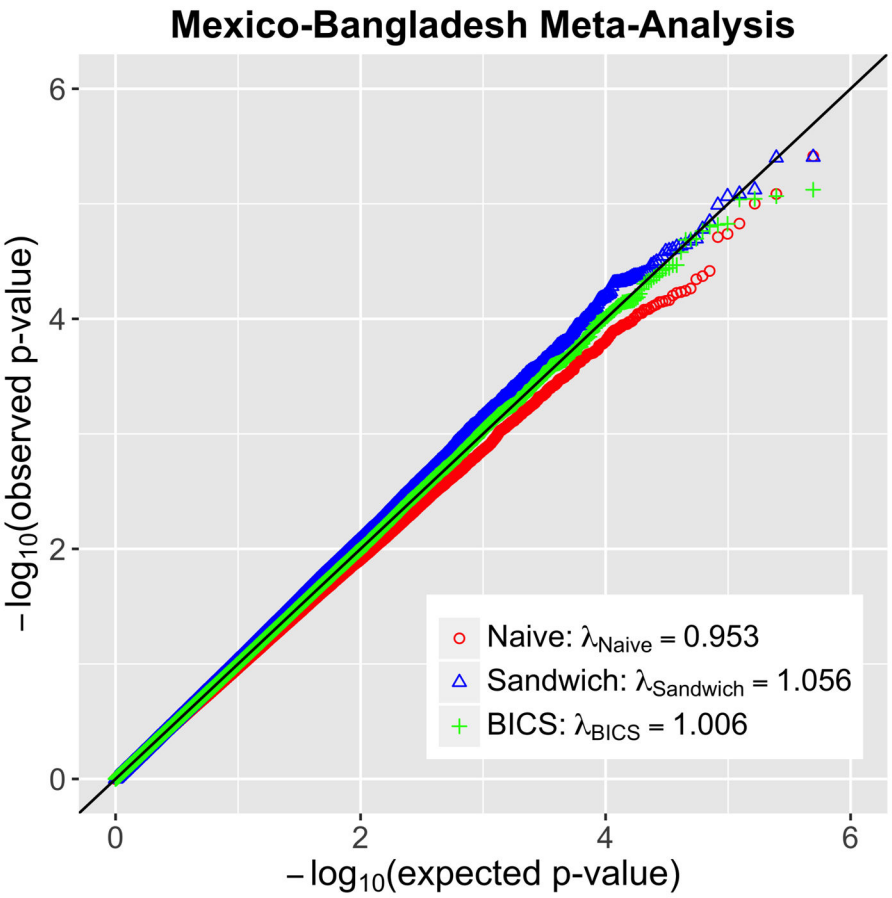

Figure 3.

The meta-analysis p-values provide our final measure of association for each interaction term. For each method of inference (naive, sandwich, BICS), we use METAL to combine the cohort-specific p-values with sample size weights. The default genomic control option in METAL is turned off. Again we see the p-values calculated using the sandwich and naive variance estimates depart from the 45-degree line very early. The BICS p-values lie almost perfectly on the line.

Table 1

Type I error rate at level $\alpha = 0.05$ when testing $H_0 : \alpha_I = 0$ with naive model-based standard error (N), sandwich standard error (S), and Bootstrap Inference with Corrected Sandwich (BICS). Simulation parameters A-E correspond to continuous outcomes and F-J to binary outcomes. Inference with the sandwich estimator shows inflated Type I error rates for continuous outcomes. Inference with the naive estimator shows inflated Type I error rates for continuous outcomes and deflated Type I error rates for binary outcomes. The BICS procedure protects the Type I error rate in all simulations.

Simulation (n)	Continuous Outcome			Binary Outcome		
	N	S	BICS	N	S	BICS
A/F						
(400)	0.073	0.063	0.052	0.037	0.047	0.050
(800)	0.077	0.060	0.053	0.039	0.048	0.051
(1600)	0.074	0.053	0.049	0.038	0.049	0.051
(3200)	0.077	0.052	0.051	0.039	0.050	0.051
B/G						
(400)	0.086	0.065	0.052	0.037	0.049	0.052
(800)	0.088	0.058	0.050	0.037	0.048	0.050
(1600)	0.089	0.055	0.050	0.037	0.050	0.051
(3200)	0.064	0.052	0.050	0.037	0.050	0.052
C/H						
(400)	0.051	0.059	0.050	0.049	0.050	0.053
(800)	0.052	0.055	0.050	0.049	0.050	0.052
(1600)	0.054	0.053	0.051	0.050	0.051	0.052
(3200)	0.060	0.052	0.051	0.049	0.050	0.051
D/I						
(400)	0.065	0.063	0.052	0.048	0.049	0.052
(800)	0.064	0.056	0.050	0.049	0.050	0.052
(1600)	0.065	0.055	0.051	0.048	0.048	0.050
(3200)	0.056	0.051	0.050	0.049	0.049	0.050
E/J						
(400)	0.074	0.065	0.053	0.035	0.048	0.051

Simulation (n)	Continuous Outcome			Binary Outcome		
	N	S	BICS	N	S	BICS
(800)	0.073	0.056	0.049	0.037	0.049	0.051
(1600)	0.076	0.054	0.051	0.036	0.049	0.051
(3200)	0.076	0.052	0.050	0.036	0.049	0.051