Tests for Gene-Environment Interactions and Joint Effects with Exposure Misclassification

Running head: GxE Interactions with Exposure Misclassification

PHILIP S. BOONSTRA, BHRAMAR MUKHERJEE*, STEPHEN B. GRUBER, JAEIL AHN, STEPHANIE L. SCHMIT, NILANJAN CHATTERJEE.

* Correspondence to Dr. Bhramar Mukherjee, Department of Biostatistics, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029, (e-mail: bhramar{at}umich.edu).

ABSTRACT

The number of methods for genome-wide testing of gene-environment interactions (GEI) continues to increase with the hope of discovering new genetic risk factors and obtaining insight into the disease-gene-environment relationship. The relative performance of these methods based on family-wise type 1 error rate and power depends on underlying disease-gene-environment associations, estimates of which may be biased in the presence of exposure misclassification. This simulation study expands on a previously published simulation study of methods for detecting GEI by evaluating the impact of exposure misclassification. We consider seven single step and modular screening methods for identifying GEI at a genome-wide level and seven joint tests for genetic association and GEI, for which the goal is to discover new genetic susceptibility loci by leveraging GEI when present. In terms of statistical power, modular methods that screen based on the marginal disease-gene relationship are more robust to exposure misclassification. Joints tests that include main/marginal effects of a gene display a similar robustness, confirming results from earlier studies. Our results offer an increased understanding of the strengths and limitations of methods for genome-wide search for GEI and joint tests in presence of exposure misclassification.
KEY WORDS: case-control; genome-wide association; gene discovery, gene-environment independence; modular methods; multiple testing; screening test; weighted hypothesis test.

Abbreviations: CC, case-control; CC(EXP), CC in the exposed subgroup; CO, case-only; CT, cocktail; DF, degree of freedom; D-G, disease-gene; EB, empirical Bayes; EB(EXP), EB in the exposed subgroup; EDGxE, joint marginal/association screening; FWER, family-wise error rate; G-E, gene-environment; GEI, gene-environment interaction; GEWIS, Gene Environment Wide Interaction Study; H2, hybrid two-step; LR, likelihood ratio; MA, marginal; OR, odds ratio; SE, sensitivity; SP, specificity; TS, two-step gene-environment screening;

Introduction

Many complex diseases (D) have a multifactorial etiology resulting from the interplay of genetic factors (G) and environmental exposures (E). Numerous statistical and epidemiological papers have considered the discovery and characterization of gene-environment interaction (GEI) [1]–[16] including discussions regarding efficiently testing GEIs [17] and conducting Gene Environment Wide Interaction Studies (GEWIS) [18, 19]. These have examined the effect of violations to gene-environment (G-E) independence in great detail.

In this paper, we build upon Mukherjee, et al. [12], who compared via simulation study the false positive rate and empirical power of several GEI search methods. We extend the simulation study in two ways. First, we augment the catalog of GEI search strategies with recently-introduced methods. Our catalog contains single-step and modular GEI search strategies, the latter of which screen for G-E and/or marginal disease-gene (D-G) association before subsequent GEI testing. Beyond these, we also evaluate "gene discovery" tests for the joint effect of gene and GEI [20, 21, 22]. These 2-degrees-of-freedom (DF) methods are less powerful than a pure marginal D-G test when there is no multiplicative GEI and empirically noted to be more powerful given modest-to-strong GEI. Power for testing the GEI component may be further increased relative to the standard 2-DF likelihood ratio (LR) test [20] through data-adaptive use of the G-E independence assumption [2, 21]. In all, we evaluate fourteen GEI and gene discovery methods.

The second extension of this paper relative to Mukherjee, et al. [12] is an evaluation of the effects of exposure misclassification on all methods. Previous studies have investigated exposure misclassification [20],[23]–[26], but no systematic published comparison under uniform simulation settings is available. Exposure misclassification/measurement error may arise in case-control studies due to recall bias, with the extent of misclassification possibly differing between cases and controls [25]–[27]. This may be particularly challenging in meta-analyses of GEI, in which the degree of measurement error in exposure data may differ across studies, leading to spurious null and non-null findings.

Misclassification in E introduces bias in the estimation of main effects and GEIs [28]–[30] and non-differential misclassification typically reduces power [31, 32]. Lindström et al. [24] study the effects of non-differential misclassification on four tests for G or GEI and found that tests with a marginal D-G association component were more robust to exposure misclassification. Recent workshops initiated by the National Institutes of Health discussed the detrimental effects of exposure misclassification, both in increased type I error and decreased power [33, 34]. Zhang et al. [23] correct the maximum likelihood estimate of odds ratios (ORs) under misclassification, using an estimate of the misclassification error rate from separate validation data. In many GEWIS, no validation data are available to implement regression calibration or other methods of adjustment from the measurement error literature [35, 36]. Stenzel et al. [37] compare several single-step procedures for GEI under the dual scenario of exposure-biased sampling and exposure misclassification. Others have studied the effect of model violations on estimation of GEI, including misspecification of the main effects in characterizing the outcome-exposure relationship [38] and the impact of unmeasured exposure confounders on GEI [22]. However, limited literature is available on studying gene-environment correlation and exposure misclassification simultaneously.

This report is organized as follows. In "Materials and Methods", we describe the testing procedures evaluated, divided into single-step or modular GEI methods and gene discovery methods. In "Simulation Settings", we describe our simulation design to evaluate each method including our approach for generating misclassified exposure data. We present operating characteristics of the methods under correctly classified and misclassified exposure scenarios in the "Results" section and conclude the paper with the "Discussion" section.

Materials and Methods

We consider a case-control study with n_1 cases and n_0 controls evaluating a set of M binary genetic markers G and a single environmental exposure E. Let $E = 1$ ($E = 0$) denote an exposed (unexposed) individual and, for each genetic marker, $G = 1$ ($G = 0$) denote whether an individual

is a carrier (non-carrier). Let D denote disease status, where $D = 1$ ($D = 0$) indicates an affected (unaffected) individual. The population parameters for a given marker are $p_{dge} \equiv \Pr(G = g, E = e|D = d)$, $d, g, e \in \{0, 1\}$. Due to the sampling mechanism, $\sum_{g,e} p_{0ge} = \sum_{g,e} p_{1ge} = 1$, and thus the corresponding frequencies follow a multinomial distribution. Table 1 defines seven log-ORs pertaining to these probabilities. The quantities θ_{GE} and γ_{GE} give G-E association in the control and case populations, respectively, α_G and α_E give marginal D-G and disease-environment association, respectively, and β_G and β_E give the respective main effects of G and E (D-G association in the sub-group $E = 0$ and disease-environment association in the sub-group $G = 0$). A non-zero value of β_{GE}, in the final row of Table 1, defines a multiplicative GEI. In its simplest form, a GEWIS tests M potential GEIs, namely $\beta_{GE} = 0$ corresponding to each marker.

Single Step Exhaustive Methods

The methods herein test all M markers for GEI, with no initial screening or prioritizing. A common adjustment to the significance threshold α_{test} is the Bonferroni correction. Each marker is tested at significance threshold α_{test}/M, controlling the family-wise error rate (FWER) at level α_{test}.

I. CASE-CONTROL (CC): The standard approach, CC calculates $\hat{\beta}_{GE}$, the maximum likelihood estimate of β_{GE} and tests $H_0 : \beta_{GE} = 0$ via Wald or LR tests using logistic regression for $P(D = 1|G, E)$.

II. CASE-ONLY (CO): Proposed by Piegorsch et al. [1], CO tests for G-E association among cases ($D = 1$), namely $H_0 : \gamma_{GE} = 0$. This can be achieved through modeling $P(G = 1|E, D = 1)$ via logistic regression. Making a rare disease approximation and assuming G-E independence in the entire population, the likelihood-ratio test for $H_0 : \gamma_{GE} = 0$ is also a valid test for $H_0 : \beta_{GE} = 0$. This does not estimate main effects of G or E (β_G or β_E).

III. EMPIRICAL BAYES (EB): To trade off between the more efficient but potentially-biased CO analysis and the always-unbiased but less efficient CC analysis, Mukherjee and Chatterjee [2] propose a shrinkage estimator based on the retrospective likelihood framework of Chatterjee and Car-

roll [39]. The estimator is given by $(\hat{w})\hat{\gamma}_{GE}+(1-\hat{w})\hat{\beta}_{GE}$, where the weight $\hat{w} = \hat{\text{Var}}(\hat{\beta}_{GE})/[\hat{\text{Var}}(\hat{\beta}_{GE})+(\hat{\beta}_{GE}-\hat{\gamma}_{GE})^2]$ adaptively controls the contribution from $\hat{\gamma}_{GE}$. The Delta method approximates the variance of this shrinkage estimator, and Wald tests based on asymptotic normality allow for inference. Regression versions of CO and EB using the retrospective likelihood framework [39] based on case-control data that provide estimates of all model parameters – not just β_{GE} – are implemented in the R package CGEN [40, 41].

Modular Methods

These methods introduce a screening or prioritizing step based on G-E or marginal D-G association before proceeding to the final GEI test. In contrast to single-step exhaustive methods, these either test only a subset of markers or vary the significance threshold for each marker based on the screening results. Statistical independence of the screening step and the final GEI test underlies these modular methods, thereby maintaining overall type 1 error.

IV. TWO-STEP G-E SCREENING (TS): Murcray et al. [4] propose this two-step procedure to leverage the efficiency of CO while maintaining robustness to G-E association:

1. Screening step: Conduct a likelihood ratio test of G-E association in the *combined* sample of cases and controls. The subset of m markers exceeding a screening significance threshold with marker-level error rate α_{scr} proceeds to the next testing step.

2. Testing step: For these m markers, conduct a CC analysis of $H_0 : \beta_{GE} = 0$ using significance threshold α_{test}/m.

Under G-E independence in the underlying population, G-E correlation in the case-enriched case-control sample indicates the presence of GEI. Screening based on γ_{GE} alone, i.e. CO, would not be asymptotically independent of the second step test statistic given by CC. The power of TS is increased when many null markers are screened out, i.e. $m \ll M$, with the magnitude of increase depending on the choice of α_{scr}. Murcray et al. [4] use $\alpha_{\text{scr}} = 0.05$, but follow-up

empirical studies show that the power increase is maximized when α_{scr} is chosen based on the case-control ratio, number of markers, and disease prevalence [11, 18]. A more recent approach from Wason and Dudbridge [13] screens based on a linear combination of the observed G-E associations resembling EB: $\hat{\gamma}_{GE} + (\hat{k})\hat{\theta}_{GE}$, where $\hat{k} = \hat{\text{Var}}(\hat{\gamma}_{GE})/\hat{\text{Var}}(\hat{\theta}_{GE})$ ensures asymptotic independence with the subsequent testing step. Like Wason and Dudbridge, we found this method to have very similar performance to TS and thus refrain from presenting the results.

V. HYBRID TWO-STEP (H2): Murcray et al. [11] later extend TS to two screening steps, one for G-E association, as in TS, and the other for marginal D-G association using α_G, the rationale being that the presence of GEIs will lead to G-E or D-G association in the case-control sample. Given a significance threshold α_{scr}, m_{assc} and m_{marg} markers, respectively, will pass each screening step, and only these are eligible for the final step GEI test. As with TS, many markers will fail both screenings, and so a less restrictive Bonferroni correction is needed at the second step CC test for GEI. The desired FWER, α_{test}, is spent between the markers from each screening step based on a pre-selected weight $\rho \in (0, 1)$. For those markers that pass both screening steps, the significance threshold is $\max\{\rho\alpha_{\text{test}}/m_{\text{marg}}, (1-\rho)\alpha_{\text{test}}/m_{\text{assc}}\}$. For the "$D$-$G$ only" markers, the significance level is $\rho\alpha_{\text{test}}/m_{\text{marg}}$, and for the "$G$-$E$ only" markers, it is $(1-\rho)\alpha_{\text{test}}/m_{\text{assc}}$. Using $\rho = 0$ makes H2 and TS equivalent. The G-E and D-G screening components are asymptotically independent of the testing step [4, 42], implying that the hybrid screening and GEI testing steps are independent. Thus, a FWER of α_{test} is maintained.

VI. COCKTAIL (CT): Hsu, et al. [14] characterize TS and H2 as special cases of a class of modular methods for GEWIS testing, comprised of separate choices of (i) screening, (ii) GEI test and (iii) type I error control modules, and propose the comprehensive class of "cocktail" procedures. In the screening step (the first module), CT adaptively tests for G-E association or marginal D-G association as in H2. In the second module, if marginal D-G association is declared statistically significant, then EB, which is independent of the D-G test, is used to test for GEI. Otherwise, CC is used, being independent of a test for G-E association in the combined case-control sample.

In the third module, and in contrast to TS and H2, no markers "fail" the screening step in CT. Rather, following the weighted hypothesis testing approach of Ionita Laza, et al. [43], α_{test} is spent differentially between all markers: those that are more significant at the screening step are given a lower significance threshold to pass at the final interaction test, as explained below.

For each marker, p^{GE} and p^{DG} denote, respectively, the p-values corresponding to the G-E and D-G screening steps. The screening module p-value is $p^{CT}_{\text{scr}} = p^{DG} I(p^{DG} \leq t) + p^{GE} I(p^{DG} > t)$, where t is a pre-specified threshold, e.g. $t = 0.001$, and $I(\cdot)$ is the indicator function. The GEI-test p-value is $p^{CT}_{\text{test}} = p^{EB} I(p^{DG} \leq t) + p^{CC} I(p^{DG} > t)$, where p^{EB} and p^{CC} are the p-values from EB and CC, respectively. To combine these modules, CT spends α_{test} between markers, comparing each p^{CT}_{test} to a potentially different significance threshold. The five markers with the smallest values of p^{CT}_{scr} have the most liberal significance threshold for testing for interaction: $\alpha_{\text{test}}/(2\times 5)$. The next 10 markers have a stricter threshold, $\alpha_{\text{test}}/(2^2 \times 10)$, and so forth. Each time, the size of the group doubles $(5, 10, 20, \ldots)$ and half of the remaining significance level $(\alpha_{\text{test}}/2, \alpha_{\text{test}}/2^2, \alpha_{\text{test}}/2^3, \ldots)$ is equally distributed to all markers in the group. The p-values p^{CT}_{scr} and p^{CT}_{test} are independent [14] but depend on a subjective threshold t. Hsu, et al. [14] proposed a modified version not requiring a threshold but for which the screening and test p-values may be correlated. Because the modified CT did not appreciably differ from CT in our simulation studies, we do not consider it further.

VII. JOINT MARGINAL/ASSOCIATION SCREENING (EDGXE): Gauderman, et al. [16] propose adding the asymptotically independent LR test statistics from the G-E and D-G screening steps and comparing to a χ^2_2 distribution as a single screening statistic. This screening step can remove markers from the GEI step, as in TS or H2, or preferentially rank markers, as in CT. We consider the latter, which had better performance in Gauderman, et al. Ege and Strachan [15] propose a similar extension: G-E and D-G associations are separately estimated for each exposure group, and the LR statistics are averaged between exposure groups. Owing to its similarity, we do not evaluate this approach.

Joint tests for discovering new loci by leveraging GEI

Even though some previously-described methods leverage information regarding G-E and/or marginal D-G association to screen markers, the final underlying null hypothesis tested is $H_0 : \beta_{GE} = 0$, and the search is one for pure GEIs. In contrast, the proceeding four strategies expand this null hypothesis and represent an agnostic search for discovery of loci, identifying those for which $\alpha_G \neq 0$, $\beta_G \neq 0$, or $\beta_{GE} \neq 0$. This modifies the definition of type I error and power relative to the standard GEI null hypothesis and results in increased rejection rates.

VIII. MARGINAL ASSOCIATION (MA): This is the standard genome-wide association study test of $H_0 : \alpha_G = 0$, the marginal D-G association test H2, CT, and EDGxE use for screening/prioritizing candidate markers. Although counter-intuitive, it is possible that $\alpha_G \neq 0$ and $\beta_G = \beta_{GE} = 0$, i.e. there is a marginal effect of G but no effect in either of the exposure sub-groups. This will hold if $\beta_E \neq 0$ and $\theta_{GE} \neq 0$ (cf. Equation W1, Web Appendix 1). Thus, due to non-linearity of the OR measures, MA may identify markers that are not associated with D in either exposure subgroup.

IX. 2-DF JOINT TESTS (JOINT(CC), JOINT(EB)) Kraft, et al. [20] suggest a joint test of $H_0 : \beta_G = \beta_{GE} = 0$, which tests for an effect of G in either exposure subgroup using standard prospective logistic regression and case-control data. We call this test JOINT(CC). A LR test statistic is compared to a χ^2_2 distribution. Rejecting H_0 does not indicate in which subgroup D-G association holds. In contrast, CC tests for a difference in association between exposure groups: $H_0 : \beta_{GE} = (\beta_G + \beta_{GE}) - \beta_G = 0$. When estimates of β_G and β_{GE} are negatively correlated, JOINT(CC) may have a larger rejection rate than CC, even when $\beta_G = 0$ [cf. page 114, 20]. We may also use the retrospective likelihood framework [39] to derive 2-DF tests for $H_0 : \beta_G = \beta_{GE} = 0$. When based on the constrained maximum likelihood, it is susceptible to bias and type 1 error inflation, like CO. Thus, we consider the EB version of this joint test that adaptively leverages G-E independence. Implemented in CGEN, this is denoted by JOINT(EB).

X. 2-DF MARGINAL+GEI TESTS (MA+CC, MA+EB) Dai, et al. [42] prove that the maximum likelihood estimate of α_G is asymptotically independent of that of both β_{GE} (CC) and γ_{GE} (CO),

and, consequently, any weighted average of the two (EB). Based on this, in a contemporaneous paper by the same authors, Dai, et al. [21] propose a simultaneous test of $H_0 : \alpha_G = \beta_{GE} = 0$. The marginal effect, α_G, is estimated via maximum likelihood, and CC, CO, or EB can estimate β_{GE}. Denoted MA+CC or MA+EB, this leverages the G-E independence assumption, leading to a more powerful test for the GEI component β_{GE} than JOINT. As with MA, these 2-DF tests may have larger rejection rates than either CC or JOINT, because α_G may be nonzero, even if $\beta_G = \beta_{GE} = 0$.

REMARK 1: The difference between JOINT(CC)/JOINT(EB) and MA+CC/MA+EB is whether one is testing the main or marginal effect of G (β_G or α_G, respectively). In the case of crossover interactions with opposite effects of G in each exposure sub-group, JOINT(CC) and JOINT(EB) are likely to be more powerful MA+CC and MA+EB.

XI. SUB-GROUP TESTS IN THE EXPOSED GROUP (CC(EXP), EB(EXP)). We propose a novel test of D-G association in the exposed group ($E = 1$) alone, namely, $H_0 : \beta_G + \beta_{GE} = 0$. This is equivalently a test of $H_0 : \beta^*_{GE} = 0$ from the constrained prospective model $\mathrm{logit}(\mathrm{P(D|G,E)}) = \beta_0 + \beta_{\mathrm{E}}\mathrm{E} + \beta^*_{\mathrm{GE}}\mathrm{G}\times\mathrm{E}$, which assumes $\beta_G = 0$. The resultant χ^2 test statistic will have one DF and be more powerful for testing pure interactions where the genetic effect is present only in the exposed group. Asymptotically, CC(EXP) is more powerful than CC if $\beta_G = 0$ [44], i.e. if the constraint is satisfied, but will lead to type 1 error when $\beta_G \neq 0$. We also use the general retrospective likelihood framework to derive a Wald test for the above hypothesis $H_0 : \beta_G + \beta_{GE} = 0$. We consider the EB version of this sub-group test in the exposed group, again using CGEN. This test, denoted by EB(EXP), adaptively leverages the G-E independence assumption.

Simulation Settings

To quantitatively evaluate these GEI methods, we modify the simulation study of Mukherjee, et al. [12], focusing on modest but plausible effect sizes for β_{GE} and α_G, based on recent published analysis findings [45]–[47]. We simulated $M = 100,000$ genetic markers with $n_0 = n_1 = 20,000$

cases and controls. Given the control prevalence of a marker G and the environmental factor E, respectively P_G and P_E, and θ_{GE}, the control probability vector $\boldsymbol{p}_0$ is obtained by solving the following system of equations:

$$\begin{aligned} \exp\{\theta_{GE}\} &= \frac{p_{000}(p_{000} - (1 - P_G - P_E))}{(1 - P_G - p_{000})(1 - P_E - p_{000})}, \\ p_{001} &= 1 - P_G - p_{000},\ p_{010} = 1 - P_E - p_{000}. \end{aligned}$$

We set $P_G = f^2 + 2f(1-f)$, where the minor allele frequency f is 0.2 for the causal marker and $f \sim \text{Unif}[0.1, 0.3]$ for null markers, and $P_E = 0.3$. For the causal marker, we used $\theta_{GE} \in \{\log(0.8), \log(1), \log(1.1)\}$, and, for the null markers, we sampled θ_{GE} from a mixture of $\text{Normal}(0, \log(1.5)/2)$, and point-mass, $\delta_0(0)$, distributions, with the proportion of zeros given by $p_{\text{ind}} \in \{0.95, 0.995, 1\}$. This is a key parameter controlling the fraction of markers correlated with E.

Choices of β_E, β_G and β_{GE}, together with $\boldsymbol{p}_0$, define the case probability vector $\boldsymbol{p}_1$ [48]: $p_{100} \propto p_{000}$, $p_{101} \propto \exp\{\beta_E\} p_{001}$, $p_{110} \propto \exp\{\beta_G\} p_{010}$, and $p_{111} \propto \exp\{\beta_E + \beta_G + \beta_{GE}\} p_{011}$. Equation W1 in Web Appendix 1 expresses the marginal log-ORs α_G and α_E as functions of $\boldsymbol{p}_0$, β_G, β_E, and β_{GE}, demonstrating that, given $\boldsymbol{p}_0$, there are three free parameters between α_G, α_E, β_G, β_E, and β_{GE}. By definition, α_E is constant across all genetic markers, i.e. for any given set of $\boldsymbol{p}_0$, β_G, β_E, and β_{GE}. However, when θ_{GE} and P_G randomly vary across markers, the strategy used by Mukherjee, et al. [12] and others, which specifies β_E, β_G and β_{GE}, will not satisfy this invariance of α_E across all markers. This incoherence is avoided by fixing $\alpha_E = 1.35$, β_G, and β_{GE}, the latter two of which are specific to each marker, and then solving for each marker-specific β_E. For the causal marker, we used $\beta_G \in \{\log(1), \log(1.2)\}$ and $\beta_{GE} < \log(1.35)$. For all other markers, we set $\beta_G = \log(1)$. Fixing α_E, β_G, and β_{GE} induces a value of α_G, the marginal genetic log-OR.

For each marker, we generate the case and control data independently from multinomial distributions using $\boldsymbol{p}_0$ and $\boldsymbol{p}_1$, respectively. To simulate exposure misclassification, we varied the sensitivity (SE) and specificity (SP) parameters. For a given marker, let $\boldsymbol{r}_1$ be the cell frequency vector for the cases. Each subject in r_{111} or r_{101}, corresponding to those for whom $E = 1$ in truth, was independently moved to r_{110} or r_{100}, respectively, with probability $1 - \text{SE}$. Simultaneously,

each subject in r_{110} or r_{100}, corresponding to $E = 0$, was moved to r_{111} or r_{101}, respectively, with probability $1 - \text{SP}$. An analogous strategy was used for the control vector $\boldsymbol{r}_0$. Perfect classification corresponds to SE $=$ SP $= 1$. We also considered non-differential misclassification, SE $=$ SP $= 0.8$, and differential misclassification, SE $= 1.0$ and SP $= 0.8$ for cases and SE $=$ SP $= 0.8$ for controls.

Web Table 1 describes additional settings: different effect or sample sizes, a rare exposure with more severe misclassification, or some null markers having non-null genetic main effects, with the results plotted in Web Figures 1–9. We generated 5,000 case-control datasets for each setting, calculating FWER, nominally 0.05, expected number of false positives, and power. We used $\alpha_{\text{scr}} = 5 \times 10^{-4}$ (TS and H2), $\rho = 0.5$ (H2), and $t = 10^{-3}$ (CT).

Results

Methods for GEI search

Table 2 presents FWER and expected number of false positives for all GEI methods. Due to differences in the null hypotheses, no such table can be meaningfully extracted for the gene discovery methods. All methods have inflated error rates under differential misclassification when $p_{\text{ind}} = 0.95$, i.e. when 5% of markers are associated with exposure, including the robust CC, identifying 3 null markers per dataset. In contrast, when all markers are independent of E ($p_{\text{ind}} = 1$), FWER is generally controlled. Under non-differential misclassification, FWER is less inflated, with the exception of CO: when $p_{\text{ind}} = 0.995$, FWER is 0.06-0.08 for EB, TS, and H2 and 0.13 for EDGxE and CT. Under perfect classification, the expected number of false positives is 2234 for CO when $p_{\text{ind}} = 0.95$. However, misclassification attenuates both G-E association and the observed GEI, and the expected number of false positives correspondingly decreases, e.g. to 1039. For EB, the adaptive linear combination of CC and CO, FWER is as large as 0.49 under differential misclassification and $p_{\text{ind}} = 0.95$.

Figure 1 plots power for the GEI methods and, for comparison, MA, against $\exp\{\beta_{GE}\}$ for

$\beta_G = \log(1.2)$, $P_E = 0.3$ and $p_{\text{ind}} = 0.995$. Web Figures 1–6 plot power under additional settings. The gene discovery method MA is considerably more powerful than the GEI methods, since α_G is typically much larger than β_{GE} in this parameterization (cf. Equation W1 in Web Appendix 1). Screening for D-G association confers robustness to misclassification, which is most evident when $\theta_{GE} = \log(0.8)$ (left column of Figure 1), but no single method dominates in all settings. Most robust to misclassification are CT and EDGxE, which use a weighted p-value screening step; H2, for which screening is a dichotomous step, also has high power but is more susceptible to misclassification. When $\theta_{GE} = \log(1)$ (middle column of Figure 1) and $\exp\{\beta_{GE}\} = 1.25$, the relative power loss of CT, EDGxE, and H2 between correct classification and non-differential misclassification is 20%, 42%, and 64%, respectively. Finally, the rejection rate of CO, which is non-monotonic with β_{GE} when $\theta_{GE} = \log(0.8)$, is explained by noting that $\gamma_{GE} = \beta_{GE} + \theta_{GE}$ (cf. Table 1).

Joint tests for discovery of new loci

Figure 2 presents the empirical rejection rates of the gene discovery methods and, for comparison, CC, against $\exp\{\beta_{GE}\}$ for $\beta_G = \log(1)$, and Web Figures 7–9 plot rejection rates under several additional settings. The rejection rate of MA is smaller than others but invariant to misclassification, as it does not depend on E; this robustness translates in part to the joint tests MA+CC and MA+EB. The data-adaptive EB methods, JOINT(EB), MA+EB, and EB(EXP), are more powerful than those maximizing the prospective likelihood alone, JOINT(CC), MA+CC, and CC(EXP) when $\theta_{GE} = 0$ or, on occasion, when misclassification attenuates the empirical θ_{GE} sufficiently to zero (bottom-right panel, Figure 2). Finally, we note that if $\beta_G \neq \log(1)$, CC(EXP) and EB(EXP), which assume this equality constraint, would be less powerful. In general, the expanded null hypothesis of the gene discovery methods is more robust to exposure misclassification, as expected. A large marginal D-G association will increase the rejection rate substantially (Web Figure 8, which differs from Figure 2 by $\beta_G = \log(1.2)$). Conversely, a small marginal D-G association, in

conjunction with misclassification, will decrease the rejection rate substantially (Web Figure 9).

Discussion

Non-differential misclassification may reduce power to detect true interactions in a GEWIS setting; however, differential misclassification may increase or decrease type I error and power. Relative to testing all markers, modular procedures that leverage empirical G-E and/or D-G associations to first screen or prioritize markers may have more power to detect GEIs. In the first such two-stage procedure, which uses only G-E association [4], the power gain depends on choosing the optimal value of screening significance level, which in turn depends on the case-control ratio, number of markers, and disease prevalence [11, 18]. A suboptimal choice may result in an empirical power curve that is non-monotonic with β_{GE}, seen here and previously [12]. Later two-step procedures that also account for D-G association (H2, EDGxE, CT) do not exhibit this undesirable property.

Because D-G association is unaffected by exposure misclassification, modular methods for GEI that use D-G association for screening or prioritization were found to be more robust to exposure misclassification. That joint tests making use of D-G association are more robust to misclassified exposure has been noted previously [24], but we document and quantify this for modern modular methods for GEI. However, even for these methods, FWER inflation under the dual challenge of differential misclassification and G-E association still remains. A limitation of all modular methods is a dependence on the choice of multiple tuning parameters: α_{scr} (TS, H2), size of weighted p-value groups (CT, EDGxE), ρ (H2), t (CT).

Gene discovery methods using joint tests for genetic association and GEI fundamentally differ and may identify genetic markers with marginal effects ($\alpha_G \neq 0$) or joint effects ($\beta_G \neq 0$, $\beta_{GE} \neq 0$). An implication of this expanded null hypothesis is that, in realistic scenarios in which more genetic markers will have detectable non-null effects for a given sample size, the number of markers identified will be considerably larger than those obtained from GEI methods. One must then investigate which markers are implicated in GEI. Any metric to evaluate gene discovery meth-

ods must take into account the context of the study, specifically what types of markers are of greater importance to identify. If discovery of new loci by leveraging GEI is the goal and marginal D-G association is anticipated, then the joint tests, particularly MA+EB and JOINT(EB), are robust to modest levels of misclassification – confirming and expanding upon the results of Lindström et al. [24] – and are able to leverage G-E independence for even greater power for testing the GEI component of a joint test.

Several limitations and possible extensions of this study exist. First, we do not consider non-parametric tree-based [49] or Boolean combinatorial methods [50] or tests for additive interaction [51]. Second, we examine the impact of exposure misclassification but do not propose any remedy. Regression calibration and imputation methods accounting for measurement error are possible solutions [35]. Most require estimation of the misclassification probabilities or existence of validation data. One might incorporate exposure quality into the construction of weights in meta-analyses of multiple studies. Third, there are many possible reasons beyond exposure misclassification that GEWIS studies lack power to detect GEIs, including small sample size [52], misclassification of the genetic markers [53], or more complex multi-marker interactions [9]. A key challenge for this and previous similar simulation studies is to realistically generate the underlying genetic architecture of a trait and magnitude and number of non-null GEI. Some specific limitations include between-marker independence, the generation of G-E associations from a mixture distribution, a lack of null markers having only main genetic effects, and considering just one causal marker for empirical power estimation (in the case of GEI). Using readily available single nucleotide polymorphism simulation routines that generate realistic linkage disequilibrium structure [54, 55] and simulating effect size parameters randomly from published estimates of genetic effect size distributions [56, 57] would make our simulation study more realistic, moving away from a fixed single parameter null/causal scenario toward a continuum of plausible genetic effect sizes. This would present challenges in terms of defining alternative metrics of *average* performance rather than simple type 1 error and power. Incorporating these into simulation studies remains an important

extension of our work.

Acknowledgements

Author affiliations: Department of Biostatistics, School of Public Health, The University of Michigan, Ann Arbor, MI (Philip S. Boonstra, Bhramar Mukherjee); Department of Bioinformatics, Biostatistics and Biomathematics, Georgetown University Medical Center, Washington, DC (Jaeil Ahn), USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA (Stephen B. Gruber, Stephanie L. Schmit); Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Rockville, MD (Nilanjan Chatterjee). R code [40] for the simulation study is available at http://www-personal.umich.edu/~philb.

This work was supported by the National Institutes of Health (grants P30 CA046592 to P.S.B., R21 ES20811 to B.M., U19 CA148107 to S.B.G., S.L.S., P.S.B., and B.M., P30 CA014089 to S.B.G., S.L.S., and B.M., T32 ES013678 to S.L.S., the Intramural Research Program of the National Cancer Institute to N.C.); and the National Science Foundation (grant DMS-1406712 to B.M.).

References

[1] Piegorsch WW, Weinberg CR, Taylor J. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med*. 1994;13(2):153-162.

[2] Mukherjee B, Chatterjee N. Exploiting gene-environment independence for analysis of case-control studies: an empirical-Bayes type shrinkage estimator to trade off between bias and efficiency. *Biometrics*. 2008;64(3):685-694.

Table 1: Seven key log-ORs defined by the case-control probabilities, p_{dge}, $d, g, e \in \{0, 1\}$, for a given marker[a]

log-OR	Value	Description
θ_{GE}	$\log\left(p_{011}p_{000}/p_{001}p_{010}\right)$	G-E given $D = 0$
γ_{GE}	$\log\left(p_{111}p_{100}/p_{101}p_{110}\right)$	G-E given $D = 1$
α_G	$\log\left([p_{111} + p_{110}][p_{001} + p_{000}]/[p_{101} + p_{100}][p_{011} + p_{010}]\right)$	D-G (marginal)
α_E	$\log\left([p_{111} + p_{101}][p_{010} + p_{000}]/[p_{110} + p_{100}][p_{011} + p_{001}]\right)$	D-E (marginal)
β_G	$\log\left(p_{000}p_{110}/p_{010}p_{100}\right)$	D-G given $E = 0$ (main)
β_E	$\log\left(p_{000}p_{101}/p_{001}p_{100}\right)$	D-E given $G = 0$ (main)
β_{GE}	$\log\left(p_{001}p_{010}p_{100}p_{111}/p_{000}p_{011}p_{101}p_{110}\right)$	Multiplicative GEI

Abbreviations: GEI, gene-environment interaction; OR, odds ratio

[a] $p_{dge} \equiv \Pr(G = g, E = e|D = d)$

Table 2: Family-wise error rate (expected number of false positives) for the GEI testing procedures as P_E, p_{ind}, and misclassification of the exposure E in the cases and controls vary[a]

Cases	Controls			Method							
{SE, SP}	{SE, SP}	p_{ind}	P_E	CC	MA	CO	EB	TS	H2	EDGxE	CT
{1,1}	{1,1}	0.950	0.3	0.05(0.05)	0.55(0.80)	1.00(2234)	0.23(0.26)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)
{1,1}	{1,1}	0.995	0.3	0.05(0.05)	0.12(0.12)	1.00(223)	0.04(0.04)	0.04(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)
{1,1}	{1,1}	1.000	0.3	0.05(0.05)	0.06(0.06)	0.05(0.05)	0.02(0.02)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.03(0.03)
{0.8,0.8}	{0.8,0.8}	0.950	0.3	0.06(0.06)	0.54(0.78)	1.00(1039)	0.39(0.49)	0.09(0.10)	0.09(0.10)	0.19(0.21)	0.18(0.19)
{0.8,0.8}	{0.8,0.8}	0.995	0.3	0.04(0.05)	0.12(0.13)	1.00(104)	0.06(0.07)	0.08(0.08)	0.06(0.06)	0.13(0.14)	0.13(0.13)
{0.8,0.8}	{0.8,0.8}	1.000	0.3	0.04(0.05)	0.05(0.05)	0.05(0.05)	0.02(0.02)	0.05(0.05)	0.05(0.05)	0.05(0.06)	0.04(0.04)
{1,0.8}	{0.8,0.8}	0.950	0.3	0.95(3)	0.55(0.80)	1.00(1670)	1.00(7)	1.00(17)	1.00(16)	1.00(24)	1.00(21)
{1,0.8}	{0.8,0.8}	0.995	0.3	0.30(0.35)	0.11(0.12)	1.00(167)	0.53(0.75)	0.99(5)	0.97(4)	1.00(7)	1.00(7)
{1,0.8}	{0.8,0.8}	1.000	0.3	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.02(0.02)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.04(0.04)
{1,1}	{1,1}	0.950	0.1	0.05(0.05)	0.13(0.14)	1.00(1591)	0.49(0.68)	0.05(0.05)	0.05(0.05)	0.04(0.05)	0.04(0.05)
{1,1}	{1,1}	0.995	0.1	0.05(0.05)	0.06(0.06)	1.00(159)	0.08(0.08)	0.04(0.04)	0.05(0.05)	0.04(0.04)	0.05(0.05)
{1,1}	{1,1}	1.000	0.1	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.02(0.02)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.04(0.04)
{0.6,0.6}	{0.6,0.6}	0.950	0.1	0.05(0.05)	0.12(0.13)	0.27(0.32)	0.06(0.06)	0.07(0.07)	0.06(0.06)	0.09(0.10)	0.13(0.13)
{0.6,0.6}	{0.6,0.6}	0.995	0.1	0.05(0.05)	0.06(0.06)	0.07(0.08)	0.03(0.03)	0.05(0.05)	0.05(0.06)	0.05(0.06)	0.05(0.05)
{0.6,0.6}	{0.6,0.6}	1.000	0.1	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.02(0.02)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.04(0.04)

Abbreviations: CC, case-control; CO, case-only; CT, cocktail; EB, empirical bayes; EDGxE, joint marginal/association screening; GEI, gene-environment interaction; H2, hybrid two-step; p_{ind}, proportion of markers in which the genetic marker (G) and exposure (E) are independent; P_E, probability that $E = 1$; SE, sensitivity, or the probability that E is correctly classified when $E = 1$ in truth; SP, specificity, or the probability that E is correctly classified when $E = 0$ in truth; TS, two-step gene-environment screening

[a] 5,000 datasets with $n = 20,000$ each of cases and controls and $M = 100,000$ genetic markers were simulated, with exactly one having multiplicative GEI ($\beta_{GE} \neq 0$). The family-wise error rate is the proportion of simulated datasets with at least one significant (null) finding, with nominal value 0.05 and standard deviation due to simulation variability of 0.003, and the expected number of false positives is the average number of significant findings per simulated dataset. The marginal exposure log-OR was $\alpha_E = \log(1.5)$ ($P_E = 0.3$) or $\log(1.75)$ ($P_E = 0.1$). For each null marker, the main genetic log-OR was $\beta_G = 0$ and the carrier prevalence was $P_G = f^2 + 2f(1 - f)$, where $f \sim \text{Unif}[0.1, 0.3]$ is the minor allele frequency. The extent of exposure misclassification increases as either SE or SP decrease.

[3] Mukherjee B, Ahn J, Rennert G, et al. Testing gene-environment interaction from case-control data: a novel study of Type-1 error, power and designs. *Genet Epidemiol*. 2008;32(7):615-626.

[4] Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*. 2009;169(2):219-226.

[5] Khoury MJ, Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies–challenges and opportunities. *Am J Epidemiol*. 2009;169(2):227-230.

[6] Chatterjee N, Wacholder S. Invited commentary: efficient testing of gene-environment interaction. *Am J Epidemiol*. 2009;169(2):231-233.

[7] Li D, Conti DV. Detecting gene-environment interactions using a combined case-only and case-control approach. *Am J Epidemiol*. 2009;169(4):497-504.

[8] Gauderman WJ,Thomas DC, Murcray CE, et al. Efficient genome-wide association testing of gene-environment interaction in case-parent trios. *Am J Epidemiol*. 2010;172(1):116-22.

[9] Thomas DC. Gene-environment-wide association studies: emerging approaches. *Nat Rev Genet*. 2010;11(4):259-272.

[10] Cornelis M, Tchetgen Tchetgen E, Liang L, et al. Gene-environment interactions in genome-wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. *Am J Epidemiol*. 2011;175(3):191-202.

[11] Murcray CE, Lewinger JP, Conti DV, et al. Sample size requirements to detect gene-environment interactions in genome-wide association studies. *Genet Epidemiol*. 2011;35(3):201-210.

[12] Mukherjee B, Ahn J, Gruber SB, et al. Testing gene-environment interaction in large-scale case-control association studies: possible choices and comparisons. *Am J Epidemiol*. 2012;175(3):177-190.

[13] Wason JMS, Dudbridge F. A general framework for two-stage analysis of genome-wide association studies and its application to case-control studies. *Am J Hum Genet*. 2012;90(5):760–773.

[14] Hsu L, Jiao S, Dai JY, et al. Powerful cocktail methods for detecting genome-wide gene-environment interaction. *Genet Epidemiol* 2012;36(3):183-194.

[15] Ege MJ, Strachan DP. Comparisons of power of statistical methods for gene-environment interaction analyses. *Eur J Epidemiol*. 2013;28(10):785-797.

[16] Gauderman WJ, Zhang P, Morrison JL, et al. Finding novel genes by testing G×E interactions in a genome-wide association study. *Genet Epidemiol*. 2013;37(6):603-613.

[17] Hutter CM, Mechanic LE, Chatterjee N, et al. Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genet Epidemiol*. 2013;37(7):643-657.

[18] Thomas DC, Lewinger JP, Murcray CE, et al. Invited commentary: *GE*-Whiz! Ratcheting gene-environment studies up to the whole genome and whole exposome. *Am J Epidemiol*. 2012;175(3):203-207.

[19] Mukherjee B, Ahn J, Gruber SB, et al. Mukherjee et al. Respond to "*GE*-Whiz! Ratcheting Up Gene-Environment Studies". *Am J Epidemiol*. 2012;175(3):208-209.

[20] Kraft P, Yen YC, Stram DO, et al. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007;63(2):111-119.

[21] Dai JY, Logsdon BA, Huang Y, et al. Simultaneously testing for marginal genetic association and gene-environment interaction. *Am J Epidemiol*. 2012;176(2):164-173.

[22] Vanderweele TJ, Mukherjee B, Chen J. Sensitivity analysis for interactions under unmeasured confounding. *Stat Med*. 2012;31(22):2552–2564.

[23] Zhang L, Mukherjee B, Ghosh M, et al. Accounting for error due to misclassification of exposures in case–control studies of gene–environment interaction. *Stat Med*. 2008;27(15):2756-2783.

[24] Lindström S, Yen Y-C, Spiegelman D, et al. The impact of gene-environment dependence and misclassification in genetic association studies incorporating gene-environment interactions. *Hum Hered*. 2009;68(2):171-181.

[25] Carroll RJ, Gail MH, Lubin JH. Case-control studies with errors in covariates. *J Am Statist Assoc*. 1993;88(421):185-199.

[26] Garcia-Closas M, Thompson WD, Robins JM. Differential misclassification and the assessment of gene-environment interactions in case-control studies. *Am J Epidemiol*. 1998;147(5):426-433.

[27] Lobach I, Fan R, Carroll RJ. Genotype-based association mapping of complex diseases: gene-environment interactions with multiple genetic markers and measurement error in environmental exposures. *Genet Epidemiol*. 2010;34(8):792-802.

[28] Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology*. 2008;19(5):640-648.

[29] Prentice, RL. Empirical evaluation of gene and environment interactions: methods and potential. *J Natl Cancer Inst*. 2011;103(16): 1209-1210.

[30] Aschard, H, Lutz S, Maus B, et al. Challenges and opportunities in genome-wide environmental interaction (GWEI) studies. *Hum Genet*. 2012;131(10):1591-1613.

[31] Garcia-Closas M, Rothman N, Lubin J. Misclassification in case-control studies of gene-environment interactions: assessment of bias and sample size. *Cancer Epidemiol Biomarkers Prev*. 1999;8(12):1043-1050.

[32] Wong, MY, Day NE, Luan, JA, et al. The detection of gene-environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement?. *Int J Epidemiol*. 2003;32(1):51-57.

[33] Bookman EB, McAllister K, Gillanders E, et al. Gene-environment interplay in common complex diseases: forging an integrative model–recommendations from an NIH workshop. For the NIH G $\times$ E Interplay Workshop participants. *Genet Epidemiol*. 2011;35(4):217-225.

[34] Mechanic LE, Chen, HS, Amos CI, et al. Next generation analytic tools for large scale genetic epidemiology studies of complex diseases. *Genet Epidemiol*. 2012;36(1):22–35.

[35] Carroll RJ, Ruppert D, Stefanski LA, et al. *Measurement error in nonlinear models: A modern perspective*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC; 2006.

[36] Spiegelman D, Rosner B, Logan R. Estimation and inference for logistic regression with covariate misclassification and measurement error in main study/validation study designs. *J Am Statist Assoc*. 2000;95(449):51–61.

[37] Stenzel SL, Ahn J, Boonstra PS, et al. The impact of exposure-biased sampling designs on detection of gene-environment interactions in case-control studies with potential exposure misclassification. *Eur J Epidemiol*. 2015;30(5):413–423.

[38] Tchetgen Tchetgen EJ, Kraft P. On the robustness of tests of genetic associations incorporat-

ing gene-environment interaction when the environmental exposure is misspecified. *Epidemiology*. 2011;22(2):257–261.

[39] Chatterjee N, Carroll RJ. Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika*. 2005;92(2):399–418.

[40] R Core Team. R: A Language and Environment for Statistical Computing, version 3.1.1. http://www.R-project.org Vienna, Austria: R Foundation for Statistical Computing; 2014.

[41] Bhattacharjee S, Chatterjee N, Han S, et al. CGEN: An R package for analysis of case-control studies in genetic epidemiology, version 3.0.0; 2012.

[42] Dai JY, Kooperberg C, Leblanc M, et al. Two-stage testing procedures with independent filtering for genome-wide gene-environment interaction. *Biometrika*. 2012;99(4):929–944

[43] Ionita-Laza I, McQueen MB, Laird NM, et al. Genomewide weighted hypothesis testing in family-based association studies, with an application to a 100K scan. *Am J Hum Genet*. 2007;81(3):607–614.

[44] Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. *Int Stat Rev*. 1991;58(2):227–240.

[45] Figueiredo JC, Lewinger JP, Song C, et al. Genotype-environment interactions in microsatellite stable/microsatellite instability-low colorectal cancer: results from a genome-wide association study. *Cancer Epidemiol Biomarkers Prev*. 2011;20(5):758–766.

[46] Garcia-Closas M, Rothman N, Figueroa JD, et al. Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. *Cancer Res*. 2013;73(7):2211–2220.

[47] Hutter CM, Chang-Claude J, Slattery ML, et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. *Cancer Res*. 2012;72(8):2036–2044.

[48] Satten GA, Kupper LL. Inferences about exposure-disease associations using probability-of-exposure information. *J Am Statist Assoc*. 1993;88(421):200–208.

[49] Ritchie MD, Hahn LW, Roodi N, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet*, 2001;69(1):138-147.

[50] Kooperberg C, Ruczinski I. Identifying interacting SNPs using Monte Carlo logic regression. *Genet Epidemiol*. 2005;28(2):157-170.

[51] Vanderweele TJ. Inference for additive interaction under exposure misclassification. *Biometrika*. 2012;99(2):502–508.

[52] Dempfle A, Scherag A, Hein R, et al. Gene–environment interactions for complex traits: definitions, methodological requirements and challenges. *Eur J Hum Genet*. 2008;16(10):1164–1172.

[53] Dudbridge F, Fletcher O. Gene-environment dependence creates spurious gene-environment interaction. *Am J Hum Genet*. 2014:95(3):301–307.

[54] Li C, Li M. GWAsimulator: a rapid whole-genome simulation program. *Bioinformatics*. 2008:24(1):140–142.

[55] Su, Zhan, Jonathan Marchini, and Peter Donnelly. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics*. 2001:27(16): 2304–2305.

[56] Park JH, Wacholder S, Gail MH, et al. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nat Genet*. 2010;42(7):570-575.

[57] Chatterjee N, Wheeler B, Sampson J, et al. Projecting the performance of risk prediction

based on polygenic analyses of genome-wide association studies. *Nat Genet*. 2013;45(4):400-405.

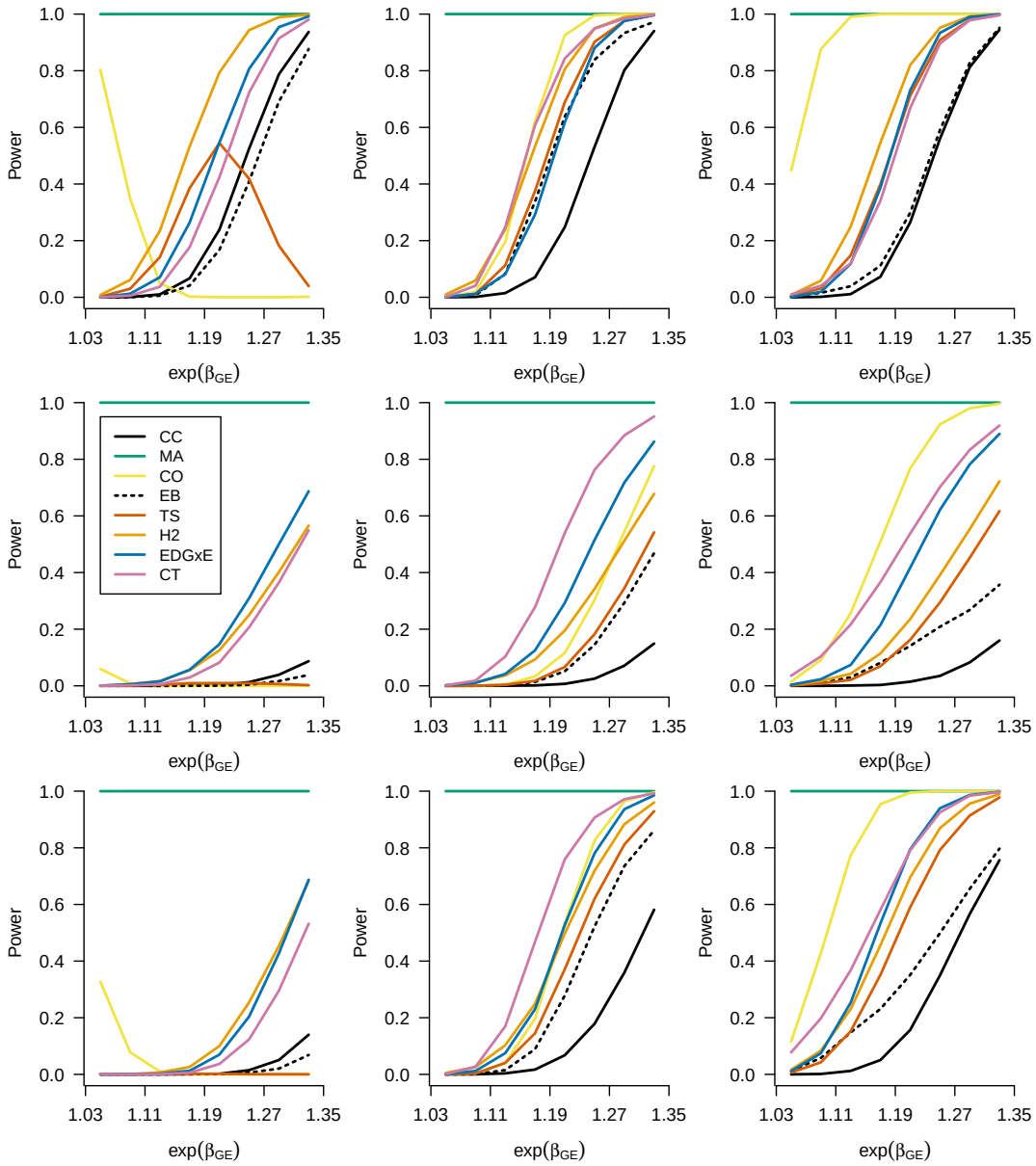

Figure 1: Empirical power to detect gene-environment interaction in one marker for 7 GEI methods (CC, case-control; CO, case-only; EB, empirical Bayes; TS, two-step gene-environment screening; H2, hybrid two-step; EDGxE, joint marginal/association screening; CT, cocktail) and the marginal (MA) method from 5,000 datasets with $n = 20,000$ each of cases and controls and $M = 100,000-1$ null genetic markers. From top to bottom, each row corresponds to perfect classification, non-differential misclassification (sensitivity and specificity of 0.8), and differential misclassification (sensitivity of 1 and specificity of 0.8 for cases and sensitivity and specificity of 0.8 for controls) of the exposure variable. From left to right, each column corresponds to $\theta_{GE} = \log(0.8)$, $\theta_{GE} = 0$, and $\theta_{GE} = \log(1.1)$. The exposure prevalence was $P_E = 0.3$ and the marginal exposure log-OR was $\alpha_E = \log(1.5)$. For the non-null marker, the main genetic log-OR was $\beta_G = \log(1.2)$ and the carrier prevalence was $P_G = 0.36$. For each null marker, $\beta_G = 0$ and $P_G = f^2 + 2f(1-f)$, where $f \sim \text{Unif}[0.1, 0.3]$ is the minor allele frequency.

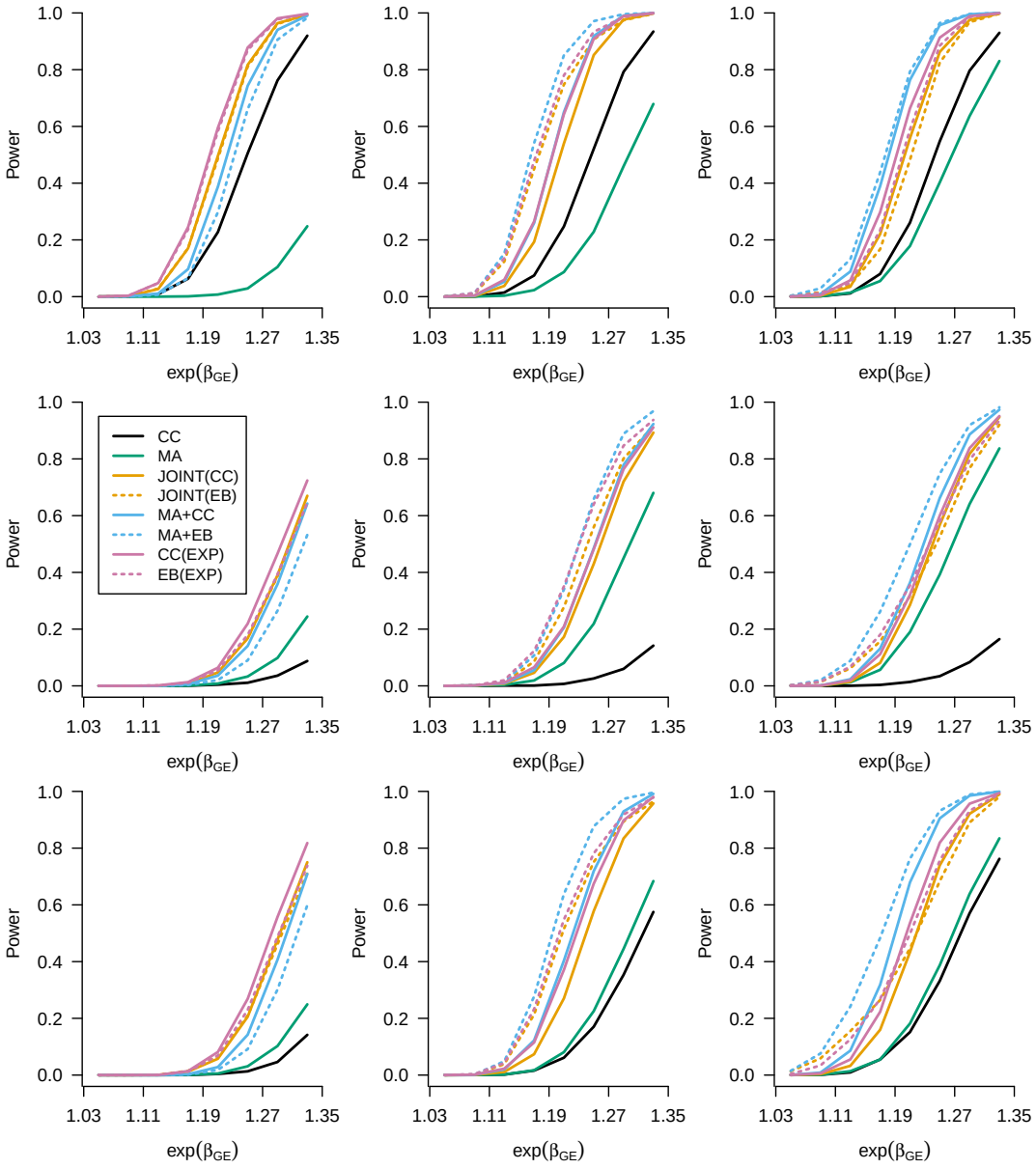

Figure 2: Empirical power for discovery of one marker for the case-control method (CC) and 7 gene discovery methods (MA, marginal; JOINT(CC), 2-DF joint test; JOINT(EB), empirical Bayes 2-DF joint test; MA+CC, marginal + CC; MA+EB, marginal + empirical Bayes; CC(EXP), CC applied to exposed subgroup; EB(EXP), empirical Bayes applied to exposed subgroup) from 5,000 datasets with $n = 20,000$ each of cases and controls. From top to bottom, each row corresponds to perfect classification, non-differential misclassification (sensitivity and specificity of 0.8), and differential misclassification (sensitivity of 1 and specificity of 0.8 for cases and sensitivity and specificity of 0.8 for controls) of the exposure variable. From left to right, each column corresponds to $\theta_{GE} = \log(0.8)$, $\theta_{GE} = 0$, and $\theta_{GE} = \log(1.1)$. The exposure prevalence was $P_E = 0.3$ and the marginal exposure log-OR was $\alpha_E = \log(1.5)$. The main genetic log-OR was $\beta_G = 0$ and the carrier prevalence was $P_G = 0.36$.