Draft Manuscript: Please do not quote, cite, or distribute

GxE Interaction Workshop: Update on the State of the Science for Analytical Methods

W. James Gauderman, Bhramar Mukherjee, Hugues Aschard, Li Hsu, Juan Pablo Lewinger, Chirag Patel, John S. Witte, Christopher Amos, Caroline Tai, David Conti, Dara Torgerson, Seunggeun Lee, Nilanjan Chatterjee

A. Introduction

Gene-environment interaction can be defined broadly as the interplay between a gene and an environmental factor as they affect some trait. For example, the epidemiologist may be interested in studying how genetic susceptibility might predispose subgroups of the population to enhanced effects of an environmental exposure. Alternatively, the geneticist may be interested in studying how exposure to an environmental fact may stimulate the expression of a gene and lead to disease.

The first paper in this series (McAllister, Mechanic, et al.) provides a broad overview of the state of the science related to gene-environment (GxE) interactions. In this paper, we summarize the latest statistical methods available for the analysis of GxE interaction. We address many of the practical questions that statisticians face as they try to uncover GxE interactions for complex traits, and summarize some of the latest approaches to address these issues (Table 1).

B. Models of GxE interaction

B.1 Basic models

Consider a study of a disease outcome (D), environmental factor (E), and genetic factor (G), with data also collected on a set of potential confounders (**C**). Exposure E can represent an exogenous environmental variable (e.g. air pollution), personal exposure (e.g. smoking), or other personal characteristic (e.g. sex). Although a single genetic locus may be of interest, most studies now genotype a large number (M, e.g. 1 million) of single nucleotide polymorphisms (SNPs) on each study subject. Each locus may be coded as G=0, 1, or 2 for the number of minor alleles, or dominant, recessive, or co-dominant coding can be used. In genome-wide association studies (GWAS), additional untyped SNP genotypes are now routinely imputed using a tool such as IMPUTE2(1) and a reference panel such as 1000 Genomes(2).

For data from a case-control study, logistic regression is typically used to model GxE interaction, with form:

$$\underline{\text{Model 1}}: \ \text{Logit}(\Pr(D=1|G,E,\mathbf{C}) = \beta_0 + \beta_G G + \beta_E E + \beta_{GxE} GxE + \boldsymbol{\beta_C}\mathbf{C}$$

Compared to an unexposed non-carrier (E=0, G=0), $\exp(\beta_G)$ measures the 'main effect' of G (E=0, G=1) and $\exp(\beta_E)$ measures the main effect of E (E=1, G=0). The corresponding odds ratio when E=1 and G=1 is $\exp(\beta_G)\exp(\beta_E)\exp(\beta_{GxE})$, and so the 'interaction odds ratio' is

$OR_{GxE} = \exp(\beta_{GxE})$ and measures the departure from the multiplicative effects of the corresponding main effects.

Model 1 builds on the typical GWAS model that does not consider GxE interaction, i.e.

$$\underline{\text{Model 2}}: \text{Logit}(\Pr(D=1|G, \mathbf{C})) = \mu_0 + \mu_G G + \boldsymbol{\mu_C}\mathbf{C}$$

where E may or may not be included in **C**. Here $OR_G = \exp(\mu_G)$ measures the 'marginal odds ratio' of G, which in the context of Model 1 can be interpreted as averaging (or marginalizing) over the exposure-specific effects of G.

For the simplest situation of a binary exposure, binary G, and no covariates, Figure S1 (online supplement) shows how the data can be represented as two 2x2 tables, and how the cell counts can be used to compute the interaction and marginal effects described above.

In the context of a cohort study, one could replace Model 1 with a log-linear model to estimate relative risks(3), or with a proportional hazards model to estimate hazard rate ratios for time-to disease data(4). For a quantitative outcome, linear regression is typically used to model main, interactive, and marginal effects (Section C.4).

B.2 Interpretations of the interaction parameters

Interpretation of GxE interaction depends on the underlying scale on which GxE effects are modeled. The classical result of Prentice and Pyke(5) ensures that estimates of (β_G, β_E, β_{GxE}, $\boldsymbol{\beta_C}$) obtained from Model 1 are valid and efficient under case-control sampling. Most commonly, interaction in epidemiology refers to the departure from multiplicative effects described above.

Another form of interaction that is not commonly assessed in practice is the departure from additivity on the absolute risk scale. Based on Model 1, the additive effect is defined as $GxE_{ADD} = \exp(\beta_G)\exp(\beta_E)\exp(\beta_{GxE}) - \exp(\beta_G) - \exp(\beta_E) + 1$. Departure from additivity implies that absolute risk-reduction associated with removal of one risk-factor depends on the levels of another and vice versa. As such, the model has direct relevance for evaluation of public health impact of risk-factor intervention(6). Further, many mechanistic forms of interactions, such as under the sufficient component cause model(6) and various modern extensions(7), have been shown to correspond to super-additive effects.

It is useful to understand the relationships between different models for interactions (Figure 1). A multiplicative GxE interaction automatically implies the effects of G and E are super-additive. Conversely, if a GxE interaction is additive, their effects are by definition sub-multiplicative. In absence of a main effect of G and/or E, the additive and multiplicative models coincide. Thus for G and/or E with weak main effects, which is often the case for common SNPs, the models may be hard to distinguish without large sample size. Recent post-GWAS studies suggest that the multiplicative model often provides reasonable approximation of GxE joint effects on disease risks(8-11). However, multiplicative models may be considered as "accepted" in traditional hypothesis testing simply due to lack of power while not necessarily being the truth. In other words, acceptance of the multiplicative null hypothesis does not automatically imply that the additive model can be rejected. Testing specifically for additive interactions in studies of breast and bladder cancer have shown that joint effects of multiple SNPs and SNPs and environmental factor are super-additive(12, 13). These studies show that investigation of interactions both scales can be informative for understanding joint G and E effects.

Mechanistic interpretation of statistical interaction is difficult because of its dependence on scale. However, certain forms of scale-invariant interactions, can provide important mechanistic insights. For example, in a "*pure interaction"* model when the effect of G is present only in the presence/absence of E (corresponding to $\beta_G=\beta_E=0$ in Model 1), interaction would be evident irrespective of scale. For example, genetic markers of acetylation in the *NAT2* gene have been associated with bladder cancer only among smokers(14). Another form of invariant interaction is the qualitative interaction where the presence of one-risk factor may reverse the effect of another, although it is arguable whether this form of interaction is biologically plausible.

C. Detecting interactions in a GWAS

C.1 Basic tests: GxE and the 2-df test

In context of Model 1, detection of multiplicative interaction is based simply on the test of the null hypothesis that $\beta_{GxE} = 0$. A Wald, Score, or likelihood ratio test may be applied. The same type of test is used if conditional logistic, log-linear, or Cox regression are used for analysis. Standard GWAS screens for marginal G effects are based on the test of H_0: $\mu_G=0$ from Model 2. A 2-degree-of-freedom procedure based on Model 1(15) or a combination of Models 1 and 2(16) can be used to test the joint null H_0: $\beta_G = \beta_{GxE} = 0$. For many models, 2-df tests have better power to detect genes than either the marginal G or 1-df GxE test alone(15).

Regardless of the specific test, one should carefully consider the inclusion of potential confounders **C**. It is also important to consider G x C and E x C interactions, as GxE interaction effects can themselves be confounded by other interactions(17). Some potential confounders seem like obvious choices for inclusion in the model, for example principal components to adjust for ancestry. Others require more judgment, such as whether or not to adjust for body mass index (BMI) in a study of gene-diet interaction. While BMI may seem like an obvious confounder, it may also be a mediator of dietary effects, and including BMI, G×BMI, and Diet×BMI as adjustments may reduce the ability to detect G×Diet signals.

In a GWAS, both G and E are each typically coded using a single trend variable, yielding a 1-df test of interaction. However, the effect of E can often be specified as categorical, ordinal or in continuous form depending on the nature of the underlying measurements. It may be desirable to code a complex exposure using a flexible model to avoid bias in the test for interaction due to model misspecification(18). This will translate to a multi-df test of interaction, which can reduce power and should be avoided if a single trend variable can be justified.

Testing for additive interaction can be numerically complex due to constraints required to guarantee risk estimates are bounded between 0 and 1 for all risk-factor combinations. For rare diseases and categorical risk factors, the additive model can be specified in the form of a general logistic regression model where the interaction term is specified by a non-linear function of the main-effects(19).

C.2 Case-only and Empirical Bayes

Piegorsch et al.(20) showed that for binary *(G, E*), the parameter β_{GxE} in Model 1 can be estimated by using data from cases only (see Figure S2). To allow for non-binary G and/or E, and adjustment for confounders, one can adopt a polytomous logistic model for case-only analysis :

Model 3: $\text{logit P}(G = g\,|E, \boldsymbol{C}, D = 1) = \gamma_0 + \gamma_1 + \gamma_{gE}E + \boldsymbol{\gamma_{gc}^T C} \qquad g = 1{,}2$

Under the assumptions of a rare disease and gene-environment (G-E) independence conditional on **C**, the likelihood ratio test for $H_0: \gamma_{1E} = \gamma_{2E} = 0$ is a valid test for interaction effects in a prospective logistic regression. Under a trend model with a single coefficient γ_E such that $\gamma_{1E} = \gamma_E$ and $\gamma_{2E} = 2\gamma_E$, the Wald test for $H_0: \gamma_E = 0$ is asymptotically equivalent to testing $H_0: \beta_{GxE} = 0$ in Model 1. A limitation of case-only analysis is that one cannot estimate the main effects β_G and β_E, and thus cannot retrieve the sub-group effects of genotype across exposure strata (e.g., $OR_{DG|E=1}$ and $OR_{DG|E=0}$). Alternative "case-only" approaches that utilize controls to estimate main effects have been proposed(21, 22).

If the G-E independence assumption is violated, estimates derived from case-only analysis will be biased(23), either toward or away from the null depending on the directions of the G-E association and GxE interaction. To reduce bias but retain some efficiency of the G-E independence assumption, Mukherjee and Chatterjee(24) proposed an empirical Bayes (EB) strategy that proceeds in a data-adaptive way. In the simple case of a binary *G* and *E,* the EB estimator $\hat{\beta}_{EB}$ is constructed using estimates from Models 1 and 3:

$$\hat{\beta}_{EB} = \hat{\gamma}_{gE} + K(\hat{\beta}_{GxE} - \hat{\gamma}_{gE}) \text{ where } K = \hat{\theta}_{gE}^2/(\hat{\sigma}_{GxE}^2 + \hat{\theta}_{gE}^2) \quad (4)$$

The intuitive explanation behind this estimate is that if there is G-E independence in the population (i.e. $\theta_{gE} = 0$), γ_{gE} and β_{GxE} will be approximately equal, and one should favor the case-only estimate $\hat{\gamma}_{gE}$ for its increased efficiency. On the other hand, if the data provide evidence of G-E dependence ($\hat{\theta}_{gE} > 0$) or if the variance of $\hat{\beta}_{GxE}$ is small ($\hat{\sigma}_{GxE}^2 \to 0$), larger weight is assigned to $\hat{\beta}_{GxE}$.

Chen et al(25) provided a more general EB framework, which is implemented in the CGEN software package(26). Li and Conti developed a Bayes model averaging (BMA) framework, where the weights in Equation 4 are defined by posterior probabilities(27). Both the EB and BMA approaches often provide greater power than a standard case-control test, while providing improved (although not perfect) control of Type 1 error compared to a case-only test in the presence of population G-E association(24). In theory, in any case-only or EB/BMA analysis, there could be bias due to G-E correlation (online supplement). Further, there could be biases in all methods due to measurement error, model mis-specification or confounding (online supplement). More empirical investigations are needed to assess the likely contribution of such bias in realistic settings. Analysis of imputed SNPs in empirical-Bayes and other retrospective methods requires special care for which methods are becoming available(28).

C.3 Efficient 2-step tests

Several '2-step' approaches have been proposed to improve the efficiency of GxE analysis while controlling Type I error, both for disease(29-34) and quantitative(35, 36) traits. All of these methods use the following general approach:

Step 1 screen: For all *M* (e.g. 1 million) SNPs, compute screening test statistic T_1 and corresponding p-value p_1.

Step 2 test: Prioritize SNPs based on p_1 (e.g. conduct step 2 only on *m* SNPs with $p_1<0.05$) and compute GxE interaction test statistic T_2 with corresponding p-value p_2. Power is increased by the need to adjust in Step 2 for only $m<<M$ tests.

A key requirement for validity of any 2-step procedure is that T_1 and T_2 must be independent. In a case-control study, two types of Step-1 screening tests have been proposed: 1) test of marginal D-G association(32), based on Model 2 and 2) test of E-G association(34), based on Model 3 applied to the combined case-control sample. The Step 2 test is based on β_{GxE} from Model 1. It has been shown that tests of $\mu_G=0$ and $\gamma_{gE}=0$ (in the combined sample) are independent of the test of $\beta_{GxE}=0$, and so either is a valid screening statistic(37). In presence of GxE interaction, one can typically expect non-zero values of both μ_G and γ_{gE}, making either the DG or EG screen useful for identifying those SNPs that are most likely to be involved in a GxE interaction. Three additional 2-step methods have been proposed including H2(33), Cocktail(31), and EDGE(30), each of which utilize DG and EG screening statistics in combination to further improve efficiency. Both standard tests and 2-step methods are implemented in the GxEScan software program(38).

C.4 Gene-environment interaction analysis for quantitative measures

Quantitative trait analyses in plant and animal models have clearly identified the importance of GxE interactions, in some cases having profound impact on phenotypes such as longevity in drosphila(39) or flowering time in arabadopsis(40). In humans, Winkler et al.(41) identified 15 loci showing evidence for gene-age dependent effects on BMI, 4 loci were not identified previously. A recent analysis suggests estimates of heritability for quantitative traits can be substantially underestimated when interaction effects are not modeled(42). For a quantitative outcome Y, a linear model of the form

$$\text{Model 5: } Y = \beta_0 + \beta_G G + \beta_E E + \beta_{GxE} GxE + \boldsymbol{\beta_C C} + \varepsilon$$

is often adopted. As for any such regression model, failure to satisfy basic model assumptions (e.g. linearity, normality of residuals ε) can lead to inflated Type I error or reduced power. As for a disease trait, GxE interaction can induce a marginal G effect, here a difference in mean Y across G. This information can be used efficiently in a 2-df joint test of G and GxE(15, 43-45), or to construct a 2-step procedure that screens on marginal-G association(36). It has also been shown that GxE interaction induces a difference in the variance of Y across G (35, 36, 46). This variance-heterogeneity information can also be used to develop valid testing procedures that are more powerful than standard tests of GxE or marginal G effects(36, 46). Standard and 2-step testing procedures are implemented in the GxEScan program(38).

C.5 Which analysis should you choose?

There have obviously been many statistical approaches developed for testing GxE interaction. Perhaps the most natural is the standard test of β_{GxE} from Model 1 or 5, as this is a simple extension of the kind of model used for marginal-G scans. However, case-only, empirical Bayes, 2-df, and 2-step procedures can offer substantial improvements in power and should be considered. For example, consider a locus G with minor allele frequency (MAF) 30%, binary E with 40% exposed, no effect of G or E alone ($OR_G = OR_E = 1.0$), and a range of possible GxE interaction effect sizes (OR_{GxE}). Figure 2 shows the number (N) of cases (assuming an equal

number of controls) required to achieve 80% power with overall Type I error rate 0.05 in genomewide GxE scan of 1 million SNPs. For example, when OR_{GxE} =1.5, a case-control study would require N=4,557 cases using a standard test of β_{GxE}=0 from Model 1. The same power can be achieved with N=2,716 using a 2-df joint test of G and GxE, N=2,160 using a case-only analysis, or with only N=1,654 using the 2-step EDGE approach. While this example is representative, the efficiency of various approaches relative to one another varies depending on the underlying true model(30, 31).

D. GxE interaction with gene sets

Set-based methods for GxE have emerged for detecting GxE effects within biologically defined sets, such as variants mapping to a particular pathway. A set-based GxE test is a single global test of interaction between an entire set of variants and the exposure of interest, rather than multiple individual tests, one per variant. The idea behind set-based methods is that accumulating multiple weak signals—possibly undetectable in isolation— across a set of variants, may result in a detectable overall GxE signal. For rare variants, set-based methods are indispensable because the power to detect GxE with any single mutation is exceedingly small.

Methods for testing set-based GxE interaction effects can be broadly classified into three categories: burden-type (BT) tests, variance component (VC) tests, and a combination of both. For BT tests, GxE testing utilizes a "G" defined as a weighted risk score computed on the gene set. The weight can be informed by DG and/or EG screening statistics (47, 48), analogous to 2-step approaches for single-SNPs. The VC approach is based on the assumption that GxE effects are random and follow an arbitrary distribution with mean 0 and variance τ^2. Testing GxE can be accomplished using a score test of H_0: τ^2= 0(49) or using a regression model of genotypic and phenotypic similarity(50, 51). BT tests perform better when many variants in the set are causal and have effects in the same direction. In contrast, VC tests are more powerful when there is heterogeneity in magnitude and direction of effects. To potentially improve power across a range of underlying scenarios, hybrid methods that combine BT and VC tests have been proposed(52-54). Set-based analogs of the 2-df joint test of G and GxE have also been developed(55, 56). The BT, VC, and hybrid methods are implemented in the MiST-I software program(54).

E. Heterogeneity in the distribution and measurement of E

An outstanding challenge in GxE analysis is the complexity of environmental data itself, which is often in stark contrast to static and discrete genetic variants(57). The analyst must consider several issues related to E when modeling and interpreting GxE interactions. First, environmental exposures are heterogeneous in their *type* (e.g., continuous or discrete). Second, different measurement modalities, for example community-level vs. home-level air pollution assessment, can differ vastly in their *measurement error* characteristics (e.g. (58, 59)), which in turn will affect bias in estimates and power to detect effects. Third, exposures and their biological effects are typically time dependent, and can vary considerably from pre-conception through adulthood. To date, GxE investigations have generally not considered how exposures change through time, and this has implications in estimation of effect sizes and interpretation. Fourth, exposures are often spatially, temporally, and/or culturally dependent. For example, air pollution levels can vary significantly between rural and urban settings and across decades(60), and racial/ethnic-specific differences in exposure to phthalates(61) and air pollution(62) have

been reported. Fifth, the correlation structure of the environmental exposures is "dense"—many environmental factors are correlated with many others (57, 63-66), making it difficult to identify the independent influence of a single exposure. These issues, and how they relate to emerging exposome-type measurements, are discussed in Patel et al. (this issue)

F. GxE analysis in a consortium setting

For standard GWAS of marginal genetic effects, achieving sufficient sample size commonly requires merging data from multiple cohorts across a consortium. Due to ethical and data protection constraints, it is usually not possible to share individual level data (to perform a so-called *mega-analysis*), and the solution has been to perform meta-analysis of cohort-specific analyses. In brief, each study performs the same analysis (e.g. application of Model 1), perhaps with some cohort-specific adjustment covariates if necessary.

Recent work has shown that meta-analysis of GxE interaction is asymptotically similar to mega-analysis(67, 68). However, there are some important considerations. First, for a binary E, a consortium may choose to perform stratified analysis. Here the goal is to estimate the marginal G effect separately within exposed and unexposed, and test for GxE interaction based on heterogeneity across E strata(e.g. (69-71)). Advantages of this stratified approach are that standard software for marginal G effects can be used, and one obtains stratified estimates and tests of G effects naturally. However, the stratified approach obviously does not extend to continuous E. Additionally, one may be tempted to over-interpret p-values of G effects within each stratum, rather than being guided by the overall test of GxE interaction that forms the basis for the primary analysis(72).

The distribution of exposure will almost certainly vary across studies in a consortium for reasons described above (Section E). For a continuous exposure, differences in distribution across cohorts are unlikely to impact the estimate of the GxE effect if the interaction effect is mostly linear (*i.e.* its direction and magnitude does not differ across the exposure range). In fact, such heterogeneity can lead to more precise estimates and greater power to detect GxE interaction (Figure S3). If GxE effects are non-linear (e.g. increased genotypic risks occur only above some threshold of ozone exposure), meta-analysis becomes more complicated and is likely to have poor power. Fundamental differences in how exposures are assessed across studies (e.g. different questionnaire items, satellite versus ground-based measurement of air pollution) may make GxE interaction for some E unanalyzable in a consortium setting.

The inclusion of diverse and/or admixed populations in a consortium may increase power to detect GxE interaction. Diverse populations can increase both genomic variation and the range of environmental exposures(73, 74). Diverse populations with varying levels of linkage disequilibrium are beneficial for fine mapping of GxE interactions to identify truly causal variation. In admixed populations, local patterns of genetic ancestry (A) can be used to perform AxE interaction analyses to increase power of discovery over traditional GxE analysis(75).

G. Why haven't many GxE interactions been identified?

While some GxE interactions have been reported (e.g. (45, 76), Ritz et al., this issue), detecting and replicating them has been a challenge and there are relatively few clear examples in the

literature. A key reason for this limited success is low statistical power, or equivalently the need for large sample sizes in order to detect interaction effects of moderate magnitude (Figure 2). Analysis (or re-analysis) in the largest possible samples (perhaps a consortium setting) using the most efficient methods may lead to the identification of additional GxE interactions.

Another issue impacting our ability to detect GxE is measurement error. For marginal G effects, the effective sample size is reduced by the LD between the analyzed and causal variants. While LD is generally high (e.g., R^2>0.8), it can vary substantially across the genome and across racial/ethnic groups. Difficulties in accurately measuring E as described above can lead to limited correlation between the observed and true E values. For example, dietary measures assessed from food frequency questionnaires commonly have low correlation with those assayed via a 24 hour dietary record (e.g., R^2<0.5). Obtaining more precise measures of E (e.g., repeat measurements, biomarkers of exposure) may be more cost-effective for improving power of GxE analysis than simply increasing sample size(77).

H. Available Software for Analysis of GxE

As is true in many areas of research, the most efficient methods of statistical analysis may not get wide use unless they are implemented in available software. For smaller scale analyses (e.g. analysis of single variants or a set of candidate genes), popular statistical software such as SAS or STATA can be used for GxE analyses. However, these programs do not scale well to genome-wide analyses or more complex models. In the sections above, we cited three software programs specifically designed for high-volume GxE analyses using novel and efficient methods(26, 38, 78). Additionally, many of the papers we cite include links to software programs that implement the corresponding methods. As we continue to move into a more high-volume, "–omics" driven research environment, it is essential that there be a strong focus on developing efficient software tools that implement evolving approaches.

I. Discussion

GxE analysis may hold the key to further understanding many complex traits. In recent years, more efficient methods for GWAS scans have been developed. These open the door to the analysis (or re-analysis) of existing resources to learn more about the range of genetic and environmental factors that affect a given trait. Modern methods for assessing exposure provide new opportunities, but also new challenges, for the detection of GxE interactions. Consortium-based studies or very large cohorts will likely be required to achieve adequate power for the analysis of GxE interaction. The study of an admixed population, either alone or as part of a consortium, may increase power for detecting GxE interaction and will certainly broaden the public-health relevance of any findings. The evolving availability of new –omics technologies will provide us with rich data resources for discovering GxE interactions and translating them into predictive/diagnostic models. Methods and software development for the analysis of GxE interaction will need to keep pace in order efficiently use these exciting new data resources.

Table 1: Challenges and potential solutions in the analysis of GxE Interactions

Challenges	Old Approach	Solutions/New Approach	For further details
Interaction can be dependent on scale	Only multiplicative scale considered	Consider evaluating interaction on both additive and multiplicative scales	Sections B and C
SNP-based analyses can lack power	Single step analysis subject to multiple comparisons burden due to large number of SNPs considered at once	Conduct more efficient 2-step tests	Section C
	Single variant approach agnostic to biological information	Conduct gene-based/set-based tests	Section D
	Individual studies report results independently	Conduct meta-analysis across studies/cohorts	Section F
	Only homogenous populations considered, typically of European decent	Consider admixture analysis, if appropriate	Section F
Exposure measurement can be inconsistent and imperfect	Individual studies independently determine method of exposure measurement	Work towards common core of exposures and definitions	Section E
	Employ easiest measurement method for largest study sample possible	Prioritize improving precision of measurements	Section G
Software is not available to conduct efficient GxE analysis	Individual analysts tweak existing software to generate limited GxE results	Implement new software designed for high-volume GxE analyses using novel methods	Section H

Figure 1: **Comparison of Additive and Multiplicative Models.** Assuming binary G and binary E, the relative risk (RR_{GE}) for joint effect, i.e relative risk for G=1 and E=1 compared to G=0 and E=0, is shown under the additive and multiplicative models. Under the two different models, the RR_{GE} is determined by underlying main effect relative risks (RR_G and RR_E), i.e. relative-risk associated with one factor (e.g G=1 vs G=0) while the other factor is fixed at baseline (e.g E=0) in two different functional forms. Throughout it is assumed that G=1 and E=1 correspond to higher risk categories compared to G=0 and E=0 ($RR_G > 1$ and $RR_E > 1$).

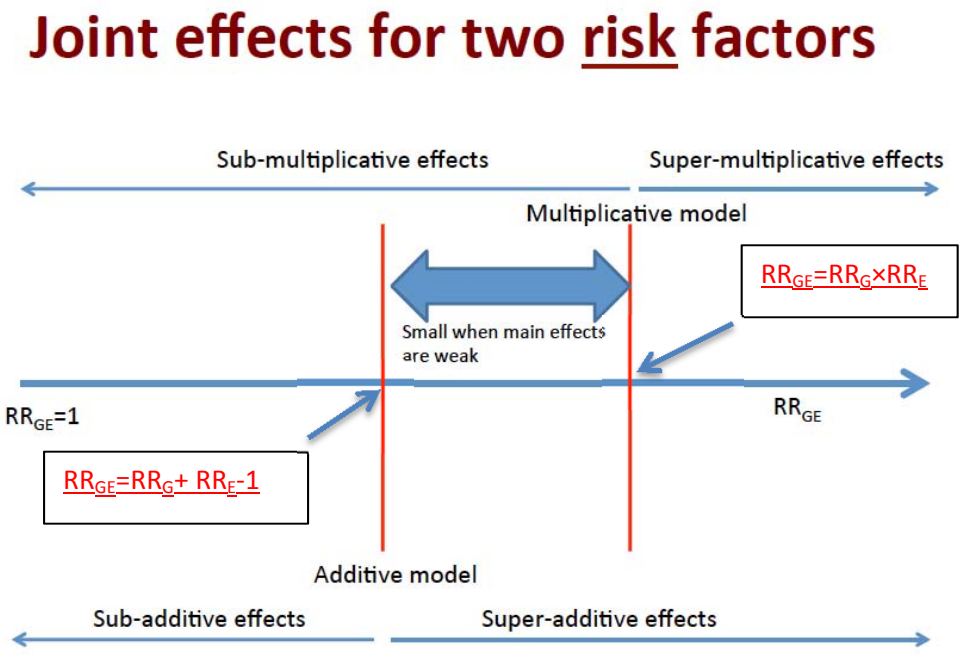
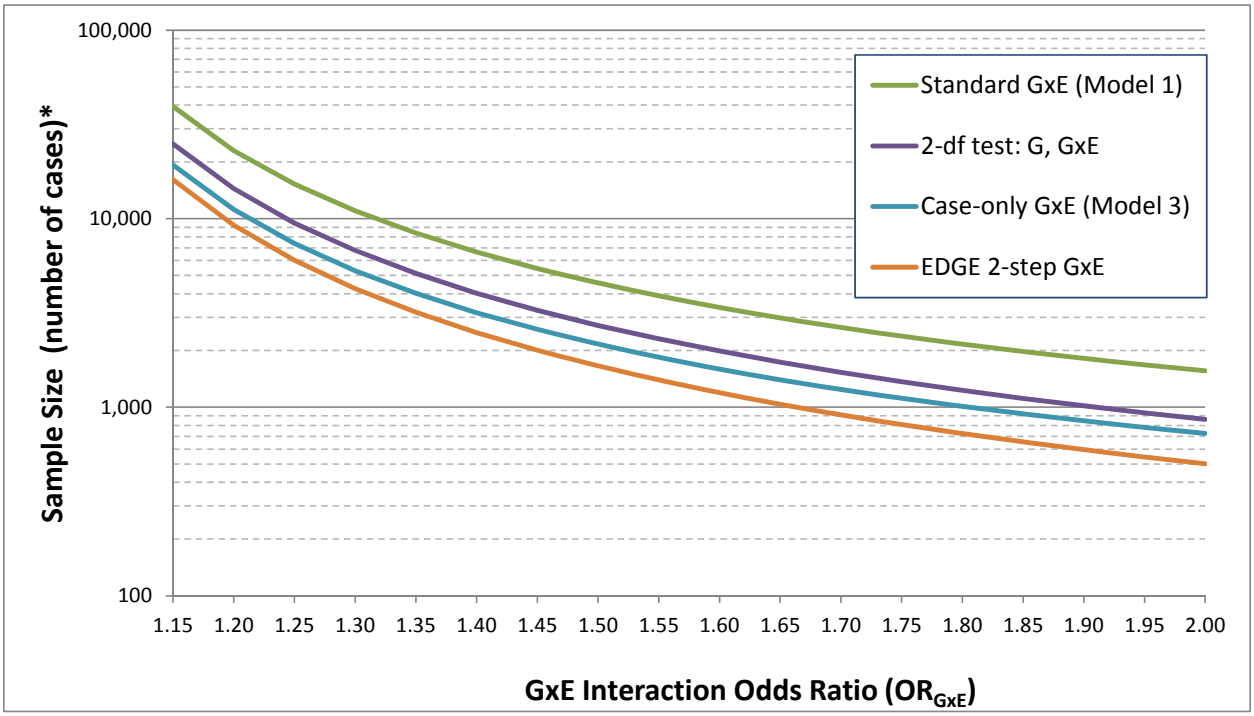


Figure 2: **Required sample size (N) versus GxE interaction effect size (OR_{GxE}) to achieve 80% power using four different analysis methods**



** N is the number of cases required to achieve 80% power, assuming an equal number of controls (except for the case-only analysis). The underlying model assumes G has minor allele frequency 30% and additive (0, 1, 2) genotype coding, E is binary with prevalence 40%, and neither G nor E has a main effect on disease risk ($\beta_G = \beta_E = 0.0$, Model 1). The calculations also assume a scan of 1 million SNPs and overall type I error rate of 0.05, yielding significance threshold for a single SNP x E interaction of 5×10^{-8}.*

References

1. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* 2010;11(7):499-511.
2. Genomes Project C, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491(7422):56-65.
3. Breslow N, Day N. *Statistical methods in cancer research: II. The design and analysis of cohort studies*. World health organization; 1987.
4. Kalbfleisch J, Prentice R. *The statistical analysis of failure time data, 2nd edition*. Indianapolis, IN: John Wiley & Sons, Inc.; 2002.
5. Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika* 1979;66:403-11.
6. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005;95 Suppl 1:S144-50.
7. VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology* 2009;20(1):6-13.
8. 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):2211-20.
9. Campa D, Kaaks R, Le Marchand L, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst* 2011;103(16):1252-63.
10. Barrdahl M, Canzian F, Joshi AD, et al. Post-GWAS gene-environment interplay in breast cancer: results from the Breast and Prostate Cancer Cohort Consortium and a meta-analysis on 79,000 women. *Hum Mol Genet* 2014;23(19):5260-70.
11. Maas P, Barrdahl M, Joshi AD, et al. Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol* 2016.
12. Garcia-Closas M, Ye Y, Rothman N, et al. A genome-wide association study of bladder cancer identifies a new susceptibility locus within SLC14A1, a urea transporter gene on chromosome 18q12.3. *Hum Mol Genet* 2011;20(21):4282-9.
13. Song M, Kraft P, Joshi AD, et al. Testing calibration of risk models at extremes of disease risk. *Biostatistics* 2015;16(1):143-54.
14. Garcia-Closas M, Malats N, Silverman D, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet* 2005;366(9486):649-59.
15. Kraft P, Yen YC, Stram DO, et al. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered* 2007;63(2):111-9.
16. Dai JY, Logsdon BA, Huang Y, et al. Simultaneously testing for marginal genetic association and gene-environment interaction. *Am J Epidemiol* 2012;176(2):164-73.
17. Keller MC. Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol Psychiatry* 2014;75(1):18-24.
18. Tchetgen Tchetgen EJ, Kraft P. On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. *Epidemiology* 2011;22(2):257-61.
19. Han SS, Rosenberg PS, Garcia-Closas M, et al. Likelihood ratio test for detecting gene (G)-environment (E) interactions under an additive risk model exploiting G-E independence for case-control data. *Am J Epidemiol* 2012;176(11):1060-7.

20. Piegorsch WW, Weinberg CR, Taylor JA. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Stat Med* 1994;13(2):153-62.
21. Umbach DM, Weinberg CR. Designing and analysing case-control studies to exploit independence of genotype and exposure. *Stat Med* 1997;16(15):1731-43.
22. Chatterjee N, Carroll RJ. Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* 2005;92:399-418.
23. Albert PS, Ratnasinghe D, Tangrea J, et al. Limitations of the case-only design for identifying gene-environment interactions. *Am J Epidemiol* 2001;154(8):687-93.
24. 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):685-94.
25. Chen YH, Chatterjee N, Carroll RJ. Shrinkage Estimators for Robust and Efficient Inference in Haplotype-Based Case-Control Studies. *J Am Stat Assoc* 2009;104(485):220-33.
26. Bhattacharjee S, Chatterjee N, Han S, et al. CGEN: An R package for analysis of case-control studies in genetic epidemiology, Version 3.8.0. http://bioconductor.org/packages/release/bioc/html/CGEN.html. 2012.
27. 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.
28. Song M, Wheeler B, Caporaso N, et al. Using imputed genotype data in the joint score-test for genetic association and gene-environment interactions in case-control studies. . *BioRxiv doi: http://dxdoiorg/101101/062075* 2016.
29. 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.
30. Gauderman WJ, Zhang P, Morrison JL, et al. Finding novel genes by testing G x E interactions in a genome-wide association study. *Genet Epidemiol* 2013;37(6):603-13.
31. Hsu L, Jiao S, Dai JY, et al. Powerful cocktail methods for detecting genome-wide gene-environment interaction. *Genet Epidemiol* 2012;36(3):183-94.
32. Kooperberg C, Leblanc M. Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet Epidemiol* 2008;32(3):255-63.
33. 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):201-10.
34. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol* 2009;169(2):219-26.
35. Pare G, Cook NR, Ridker PM, et al. On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. *PLoS Genet* 2010;6(6):e1000981.
36. Zhang P, Lewinger JP, Conti D, et al. Detecting Gene-Environment Interactions for a Quantitative Trait in a Genome-Wide Association Study. *Genet Epidemiol* 2016;40(5):394-403.
37. 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):929-44.
38. Gauderman W, Morrison J, Zhang P. GxEScan: A Program for Detecting GxE Interactions in a Genome-wide Association Study, Version 0.5.0. http://biostats.usc.edu/software. 2016.
39. Vieira C, Pasyukova EG, Zeng ZB, et al. Genotype-environment interaction for quantitative trait loci affecting life span in Drosophila melanogaster. *Genetics* 2000;154(1):213-27.
40. Juenger TE, Sen S, Stowe KA, et al. Epistasis and genotype-environment interaction for quantitative trait loci affecting flowering time in Arabidopsis thaliana. *Genetica* 2005;123(1-2):87-105.

41. Winkler TW, Justice AE, Graff M, et al. The Influence of Age and Sex on Genetic Associations with Adult Body Size and Shape: A Large-Scale Genome-Wide Interaction Study. *PLoS Genet* 2015;11(10):e1005378.
42. Zuk O, Hechter E, Sunyaev SR, et al. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* 2012;109(4):1193-8.
43. Aschard H, Hancock DB, London SJ, et al. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Hum Hered* 2011;70(4):292-300.
44. Manning AK, LaValley M, Liu CT, et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP x environment regression coefficients. *Genet Epidemiol* 2011;35(1):11-8.
45. Hancock DB, Soler Artigas M, Gharib SA, et al. Genome-wide joint meta-analysis of SNP and SNP-by-smoking interaction identifies novel loci for pulmonary function. *PLoS Genet* 2012;8(12):e1003098.
46. Aschard H, Zaitlen N, Tamimi RM, et al. A nonparametric test to detect quantitative trait loci where the phenotypic distribution differs by genotypes. *Genet Epidemiol* 2013;37(4):323-33.
47. Jiao S, Hsu L, Bezieau S, et al. SBERIA: set-based gene-environment interaction test for rare and common variants in complex diseases. *Genet Epidemiol* 2013;37(5):452-64.
48. Liu Q, Chen LS, Nicolae DL, et al. A unified set-based test with adaptive filtering for gene-environment interaction analyses. *Biometrics* 2015.
49. Lin X, Lee S, Christiani DC, et al. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics* 2013;14(4):667-81.
50. Tzeng JY, Zhang D, Pongpanich M, et al. Studying gene and gene-environment effects of uncommon and common variants on continuous traits: a marker-set approach using gene-trait similarity regression. *Am J Hum Genet* 2011;89(2):277-88.
51. Zhao G, Marceau R, Zhang D, et al. Assessing gene-environment interactions for common and rare variants with binary traits using gene-trait similarity regression. *Genetics* 2015;199(3):695-710.
52. Lin X, Lee S, Wu MC, et al. Test for rare variants by environment interactions in sequencing association studies. *Biometrics* 2015.
53. Jiao S, Peters U, Berndt S, et al. Powerful Set-Based Gene-Environment Interaction Testing Framework for Complex Diseases. *Genet Epidemiol* 2015.
54. Su Y-R, Di C, Hsu L, et al. A Unified Powerful Set-based Test for Sequencing Data Analysis of GxE Interactions. *Biostatistics* 2016, in press.
55. Kazma R, Cardin NJ, Witte JS. Does accounting for gene-environment interactions help uncover association between rare variants and complex diseases? *Hum Hered* 2012;74(3-4):205-14.
56. Broadaway KA, Duncan R, Conneely KN, et al. Kernel Approach for Modeling Interaction Effects in Genetic Association Studies of Complex Quantitative Traits. *Genet Epidemiol* 2015;39(5):366-75.
57. Ioannidis JP, Loy EY, Poulton R, et al. Researching genetic versus nongenetic determinants of disease: a comparison and proposed unification. *Sci Transl Med* 2009;1(7):7ps8.
58. Palmer CD, Lewis ME, Geraghtya CM, et al. Determination of lead, cadmium and mercury in blood for assessment of environmental exposure: A comparison between inductively coupled plasma–mass spectrometry and atomic absorption spectrometry. *Spectrochimica Acta Part B: Atomic Spectroscopy* 2006;61(8):980-90.
59. Zeger SL, Thomas D, Dominici F, et al. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect* 2000;108(5):419-26.
60. Lurmann F, Avol E, Gilliland F. Emissions reduction policies and recent trends in Southern California's ambient air quality. *J Air Waste Manag Assoc* 2015;65(3):324-35.

61. Silva MJ, Barr DB, Reidy JA, et al. Urinary levels of seven phthalate metabolites in the U.S. population from the National Health and Nutrition Examination Survey (NHANES) 1999-2000. *Environ Health Perspect* 2004;112(3):331-8.
62. Zou B, Peng F, Wan N, et al. Spatial cluster detection of air pollution exposure inequities across the United States. *PLoS One* 2014;9(3):e91917.
63. Gauderman WJ, Avol E, Gilliland F, et al. The effect of air pollution on lung development from 10 to 18 years of age. *N Engl J Med* 2004;351(11):1057-67.
64. Patel CJ, Ioannidis JP. Placing epidemiological results in the context of multiplicity and typical correlations of exposures. *J Epidemiol Community Health* 2014;68(11):1096-100.
65. Patel CJ, Manrai AK. Development of exposome correlation globes to map out environment-wide associations. *Pac Symp Biocomput* 2015:231-42.
66. Smith GD, Lawlor DA, Harbord R, et al. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med* 2007;4(12):e352.
67. Lin DY, Zeng D. Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet Epidemiol* 2010;34(1):60-6.
68. Sung YJ, Schwander K, Arnett DK, et al. An empirical comparison of meta-analysis and mega-analysis of individual participant data for identifying gene-environment interactions. *Genet Epidemiol* 2014;38(4):369-78.
69. Randall JC, Winkler TW, Kutalik Z, et al. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet* 2013;9(6):e1003500.
70. Myers RA, Scott NM, Gauderman WJ, et al. Genome-wide interaction studies reveal sex-specific asthma risk alleles. *Hum Mol Genet* 2014;23(19):5251-9.
71. Magi R, Lindgren CM, Morris AP. Meta-analysis of sex-specific genome-wide association studies. *Genet Epidemiol* 2010;34(8):846-53.
72. Aschard H. A perspective on interaction effects in genetic association studies. *Genet Epidemiol* 2016, in press.
73. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015;526(7571):68-74.
74. Moore SC, Gunter MJ, Daniel CR, et al. Common genetic variants and central adiposity among Asian-Indians. *Obesity (Silver Spring)* 2012;20(9):1902-8.
75. Schoeps A, Rudolph A, Seibold P, et al. Identification of new genetic susceptibility loci for breast cancer through consideration of gene-environment interactions. *Genet Epidemiol* 2014;38(1):84-93.
76. Lubin JH, Kogevinas M, Silverman D, et al. Evidence for an intensity-dependent interaction of NAT2 acetylation genotype and cigarette smoking in the Spanish Bladder Cancer Study. *Int J Epidemiol* 2007;36(1):236-41.
77. 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):51-7.
78. Sun J, Zheng Y, Hsu L. MiST-I: A program for set-based gene-environment interaction tests. Version 1.0. http://research.fhcrc.org/hsu/en/software.html. 2014.