RESEARCH ARTICLE

Genetic Epidemiology
OFFICIAL JOURNAL
INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY
www.geneticepi.org

Novel Likelihood Ratio Tests for Screening Gene-Gene and Gene-Environment Interactions With Unbalanced Repeated-Measures Data

Yi-An Ko,[1]* Paramita Saha-Chaudhuri,[2] Sung Kyun Park,[3] Pantel Steve Vokonas,[4,5] and Bhramar Mukherjee[1]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan; [2]Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina; [3]Department of Epidemiology, Department of Environmental Health Sciences, University of Michigan, Ann Arbor, Michigan; [4]VA Normative Aging Study, Veterans Affairs Boston Healthcare System, Boston, Massachusetts; [5]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts

Received 3 December 2012; Revised 29 March 2013; accepted revised manuscript 20 May 2013.
Published online 24 June 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/gepi.21744

ABSTRACT: There has been extensive literature on modeling gene-gene interaction (GGI) and gene-environment interaction (GEI) in case-control studies with limited literature on statistical methods for GGI and GEI in longitudinal cohort studies. We borrow ideas from the classical two-way analysis of variance literature to address the issue of robust modeling of interactions in repeated-measures studies. While classical interaction models proposed by Tukey and Mandel have interaction structures as a function of main effects, a newer class of models, additive main effects and multiplicative interaction (AMMI) models, do not have similar restrictive assumptions on the interaction structure. AMMI entails a singular value decomposition of the cell residual matrix after fitting the additive main effects and has been shown to perform well across various interaction structures. We consider these models for testing GGI and GEI from two perspectives: likelihood ratio test based on cell means and a regression-based approach using individual observations. Simulation results indicate that both approaches for AMMI models lead to valid tests in terms of maintaining the type I error rate, with the regression approach having better power properties. The performance of these models was evaluated across different interaction structures and 12 common epistasis patterns. In summary, AMMI model is robust with respect to misspecified interaction structure and is a useful screening tool for interaction even in the absence of main effects. We use the proposed methods to examine the interplay between the hemochromatosis gene and cumulative lead exposure on pulse pressure in the Normative Aging Study.
Genet Epidemiol 37:581–591, 2013. © 2013 Wiley Periodicals, Inc.

KEY WORDS: likelihood ratio test; longitudinal data; mixed model; nonadditivity; parametric bootstrap; singular value decomposition; two-step regression

Introduction

Prospective cohort studies examining gene-gene or gene-environment interaction (GGI or GEI) effects on disease-related quantitative traits have received considerable attention in recent years [Bookman et al., 2011; Fan et al., 2012]. The detection of GEI plays a critical role in identifying a subpopulation of the genetically susceptible individuals that are strongly affected by an adverse exposure. A better understanding of GEI may lead to the development of more effective disease prevention and intervention strategies. Studies of GEI in relation to disease development are facilitated by life-time characterization of exposure data, which are often available in prospective cohort studies. Repeated-measures design in a prospective cohort study may increase power to detect interaction effects [Wong et al., 2003] and provide better ways to handle exposure measurement error. In addition, repeated-measures data provide valuable information for delineating potentially time-dependent form of GGI or GEI, thereby permitting a much more detailed assessment of the dynamically evolving interplay between genes and environment.

Cohort studies for GGI or GEI are typically characterized by unequal sample size in each genotype-genotype or genotype-exposure configuration as a result of unbalanced allele frequencies and heterogeneous environmental exposure distributions in a population. A common analysis strategy for such unbalanced data involves modeling GGI or GEI by a product term in a regression setting, implying that the effect of two factors may not be purely additive in their contribution to the quantitative trait. Alternatively, one can try to model the interaction term in the generalized additive mixed model framework with nonlinear exposure and time effects [Lin and Zhang, 1999], but tests for such nonparametric, smoothed interaction terms may yield reduced power for moderate sample size. Therefore, flexible yet parsimonious modeling of GGI or GEI is of interest in the longitudinal setting. In this paper, we propose likelihood ratio

Supporting Information is available in the online issue at wileyonlinelibrary.com.

*Correspondence to: Yi-An Ko, Department of Biostatistics, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109-2029. E-mail: yianko@umich.edu

tests (LRTs) for GGI and GEI using a sparse representation of interaction borrowing ideas from the classical analysis of variance (ANOVA) literature.

Genetic factors (G) and environmental exposures (E) are frequently treated as binary or ordered categorical variables. Consequently, GGI and GEI are often analyzed in the form of a two-way table. Considering G as a row variable with I categories and E as a column variable with J categories, the mean structure of a general two-way classification model for analyzing row × column interactions is given by

$$\mu_{ij} = \mu + R_i + C_j + \gamma_{ij}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad (1)$$

where μ_{ij} is the expected (mean) value of a quantitative trait corresponding to the ith row and the jth column, μ is the grand mean, R_i is the additive main effect of the ith row, C_j is the additive main effect of the jth column, and γ_{ij} is the nonadditive effect of the ith row and the jth column. The sum-to-zero conditions,

$$\sum_i R_i = \sum_j C_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0, \quad (2)$$

ensure identifiability of the parameters in (1), so the degrees of freedom for testing γ_{ij} in a fully saturated model is $(I-1)(J-1)$. While a saturated model (1) is flexible for estimation of γ_{ij}, the degrees of freedom for interaction tests can increase considerably for finely cross-classified tables, which is inefficient and may result in low power for detecting GGI or GEI.

To improve the power of the test for GGI and GEI in longitudinal cohort studies, we explore alternative parsimonious interaction structures that were proposed in the classical ANOVA literature for testing interaction with only one observation per cell. Several models are summarized in the following:

Model (a): $\mu_{ij} = \mu + R_i + C_j + \theta R_i C_j$ [Tukey, 1949]

Model (b): $\mu_{ij} = \mu + R_i + C_j + \lambda_i C_j$ [Mandel, 1961]

Model (c): $\mu_{ij} = \mu + R_i + C_j + R_i \eta_j$ [Mandel, 1961]

Model (d): $\mu_{ij} = \mu + R_i + C_j + \theta R_i C_j + \lambda_i C_j + R_i \eta_j$ [Tukey, 1962]

with constraints $\sum_{i=1}^{I} \lambda_i = \sum_{j=1}^{J} \eta_j = 0$ for models (b) and (c), respectively, and additional constraints $\sum_{i=1}^{I} \lambda_i R_i = \sum_{j=1}^{J} \eta_j C_j = 0$ for model (d). The multiplicative interaction term is proportional to the main effects of one or both factors. The null hypotheses of no interaction for models (a)–(d) are $\theta = 0$, $\lambda_i = 0$ $(i = 1, \ldots, I-1)$, $\eta_j = 0$ $(j = 1, \ldots, J-1)$, and $\theta = \lambda_i = \eta_j = 0$ $(i = 1, \ldots, I-2; j = 1, \ldots, J-2)$, corresponding to 1, $I-1$, $J-1$, and $I+J-3$ df for the tests of interaction effects, respectively. A more flexible potential alternative is the AMMI model [Gollob, 1968; Mandel, 1971]

$$\text{Model (e)}: \quad \mu_{ij} = \mu + R_i + C_j + \sum_{m=1}^{M} d_m \alpha_{im} \beta_{jm} + \gamma^*_{ij},$$

where M represents the number of interaction factors being extracted, $M \leq \min(I-1, J-1)$, and a residual γ^*_{ij} remains if not all interaction factors are used. The terms $\{\alpha_{im}\beta_{jm}\}$ can the considered as the weights corresponding to a multiplicative contrast among $\{\gamma_{ij}\}$ with $\sum_i \alpha_{im} = \sum_j \beta_{jm} = 0$ and $\sum_i \alpha_{im}\alpha_{im'} = \sum_j \beta_{jm}\beta_{jm'} = 0$ for $m \neq m'$. For normalized contrasts ($\sum_i \alpha^2_{im} = \sum_j \beta^2_{jm} = 1$), $\{d_m, \alpha_{im}, \beta_{jm}\}$ can be obtained by applying singular value decomposition (SVD) to $\{\gamma_{ij}\}$. Since the motivation for using an AMMI model is to extract a low-rank approximation to the interaction matrix to save degrees of freedom and thus to enhance efficiency for the test, we focus on AMMI models with $M = 1$ (AMMI1). For all subsequent discussions, model (e) refers to AMMI1 model. The null hypothesis of no interaction for AMMI1 model is $H_0 : d_1 = 0$.

Models (a)–(e) were conceived from a statistical objective of reducing degrees of freedom and enhancing power of tests for interaction. They have been used in designed genotype-by-environment yield trials in agricultural studies [Crossa et al., 1990; Freeman, 1973; Zobel et al., 1988]. These models were not conceived from a mechanistic or human biological perspective. Model (a) has recently been used to test for genetic effects in case-control studies [Chatterjee et al., 2006] and repeated-measures data of complex traits [Maity et al., 2009]. Models (a)–(e) have also been applied for GGI effects on quantitative traits in cross-sectional studies [Barhdadi and Dubé, 2010]. In unbalanced designs, the sums of squares associated with the two factors and their interaction are not orthogonal to one another. Consequently, the difficulties that arise in applying these nonlinear interaction models to unbalanced data involve obtaining unbiased parameter estimates, partitioning the sums of squares, deriving the appropriate test statistics and their null distributions. Mukherjee et al. [2012] proposed a screening tool for GGI and GEI using cell means from an unbalanced repeated-measures array. This approach is appealing due to a closed form analytical expression of the test statistic. However, violations in the homoscedasticity assumption of cell-mean error distributions result in inflated type I error. While their proposed resampling-based method recognizes unbalanced, repeated-measures data structure, the test implemented for AMMI models lacks power because it was not based on a theoretically derived pivot but an ad hoc extension of the balanced, cross-sectional case.

To overcome some of the limitations of the previous methods, we propose alternate approaches to explore GGI and GEI using models (a)–(e). We first propose our improved cell-mean approach that properly handles unbalanced data. Specifically, we adapt and modify the test proposed by Boik [1989] under a reduced-rank model for application to GGI/GEI using AMMI models. Next, we extend models (a)–(e) to the repeated-measures setting using a mixed-effects modeling framework. We then develop a parametric bootstrap resampling approach by replacing the ad hoc pivot in Mukherjee et al. [2012] with a LRT-based pivot derived from the maximum likelihood (ML) under a nonlinear mixed-effects model. The power and type I error of our proposed tests are examined through a series of

simulation studies. Lastly, we apply the proposed methods to a GEI study concerning the modifying effects of polymorphisms in the hemochromatosis gene (*HFE*) on the association between cumulative lead exposure and pulse pressure (PP) [Zhang et al., 2010]. Subject- and time-specific contributions to GEI are investigated using outputs from the AMMI model.

Methods

LRT Based on Cell Means

Following the notations in (1), let y_{ijkh} be the hth measurement corresponding to the kth individual in the (i, j)th cell (or equivalently, row i and column j) in a longitudinal cohort study, $i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, N_{ij}, h = 1, \ldots, n_{ijk}$. Let N denote the total number of individuals, $N = \sum_i \sum_j N_{ij}$. Let $\bar{\boldsymbol{Y}} = \{\bar{Y}_{ij}\}$ be the $I \times J$ matrix of sample means with $\bar{Y}_{ij} = \sum_{k=1}^{N_{ij}} \sum_{h=1}^{n_{ijk}} y_{ijkh} / \sum_{k=1}^{N_{ij}} n_{ijk}$. Let $\boldsymbol{L}$ be the matrix of main effects, parameterized as $\boldsymbol{L} = \mathbf{1}_I \mu \mathbf{1}_J' + \boldsymbol{R}\mathbf{1}_J' + \mathbf{1}_I \boldsymbol{C}'$, where $\mathbf{1}_\nu$ is a length-ν vector of ones, and $\boldsymbol{R} = (R_1, \ldots, R_I)'$, and $\boldsymbol{C} = (C_1, \ldots, C_J)'$ are the parameter vectors representing row and column effects, respectively. Let $\boldsymbol{\Gamma}$ be the $I \times J$ matrix of interaction effects, so the mean structures of models (a)–(e) can be expressed as $\boldsymbol{E}(\bar{\boldsymbol{Y}}) = \boldsymbol{L} + \boldsymbol{\Gamma}$. Throughout our treatment of the problem, we consider the drop-outs in longitudinal studies to be missing at random, leading to the unbalanced data structure.

We propose to use an empirical variance estimate for the variance of $\bar{Y}_{ij}$ (denoted as δ_{ij}^2) that accounts for within-subject correlation. Let $\sigma^2 \boldsymbol{P}(\boldsymbol{\rho})$ be a symmetric $n_{ijk} \times n_{ijk}$ within-subject covariance matrix, where $\boldsymbol{\rho}$ is a $s \times 1$ parameter vector that fully characterizes the correlation matrix $\boldsymbol{P}(\boldsymbol{\rho})$, and σ^2 is a scale parameter. Both $\boldsymbol{\rho}$ and σ^2 can be estimated by Pearson residuals, namely, $\hat{r}_{ijkh} = y_{ijkh} - \bar{y}_{ijk}$ [Liang and Zeger, 1986]. The pooled estimate of σ^2 is

$$\hat{\sigma}^2 = \sum_i \sum_j (N_{ij} - 1)\hat{\sigma}_{ij}^2 \Big/ \left(\sum_i \sum_j N_{ij} - IJ \right),$$

where

$$\hat{\sigma}_{ij}^2 = \sum_{k=1}^{N_{ij}} \sum_{h=1}^{n_{ijk}} \hat{r}_{ijkh}^2 \Big/ \left(\sum_{k=1}^{N_{ij}} n_{ijk} - 1 \right).$$

The estimation of $\boldsymbol{\rho}$ is conditional on the correlation structure. For a compound symmetric correlation structure, $s = 1$, $\mathrm{corr}(y_{ijkh}, y_{ijkh'}) = \rho$ for $h \neq h'$. The pooled estimate for ρ is

$$\hat{\rho} = \sum_i \sum_j (N_{ij} - 1)\hat{\rho}_{ij} \Big/ \left(\sum_i \sum_j N_{ij} - IJ \right),$$

where

$$\hat{\rho}_{ij} = \sum_{k=1}^{N_{ij}} \sum_{h>h'} \hat{r}_{ijkh}\hat{r}_{ijkh'} \Big/ \left\{ \hat{\sigma}^2 \left[\sum_{k=1}^{N_{ij}} \frac{1}{2} n_{ijk}(n_{ijk} - 1) - 1 \right] \right\}.$$

Finally, the empirical variance estimate for $\bar{Y}_{ij}$ is given by

$$\hat{\delta}_{ij}^2 = \frac{\hat{\sigma}^2}{n_{ij}} + \frac{2\hat{\sigma}^2}{n_{ij}^2} \sum_{k=1}^{N_{ij}} \sum_{h>h'} \hat{\mathrm{corr}}(y_{ijkh}, y_{ijkh'}),$$

$$\text{where} \quad n_{ij} = \sum_{k=1}^{N_{ij}} n_{ijk}. \tag{3}$$

Given $\hat{\delta}_{ij}$, we maximize the likelihood of $\bar{\boldsymbol{Y}}$ under the normality assumption, namely, $\mathrm{vec}(\bar{\boldsymbol{Y}}) \sim \mathcal{N}\big(\mathrm{vec}(\boldsymbol{L}) + \mathrm{vec}(\boldsymbol{\Gamma}), \mathrm{Diag}(\hat{\delta}_{ij}^2)\big)$. Maximizing the log-likelihood is equivalent to least squares fitting of μ_{ij} subject to weights $1/\hat{\delta}_{ij}^2$. For classical interaction models (a)–(d) involving nonlinearity in the parameters, the ML estimates for $\boldsymbol{L}$ and $\boldsymbol{\Gamma}$ are obtained using a quasi-Newton method in R [R Core Team, 2012] with function "optim" and L-BFGS-B algorithm [Nocedal and Wright, 1999]. Quasi-Newton methods are sequential line search algorithms, and generally require only the gradient of the objective to be computed at each iterate. When convergence is reached, we calculate the log-likelihoods under the null ($\hat{\ell}_0$) and under the alternative ($\hat{\ell}_1$) to construct the LRT statistic: $-2(\hat{\ell}_0 - \hat{\ell}_1)$. Under H_0, the LRT statistic approximately follows a central chi-square distribution with df = 1, $I - 1$, $J - 1$, and $I + J - 3$ for models (a)–(d), respectively. The comparison of empirical quantiles of the LRT statistics with chi-square quantiles is presented in Supplementary Fig. S1.

Boik [1989] proposed the likelihood ratio criterion to test the rank of $\boldsymbol{\Gamma}$ for unbalanced data without repeated measures. For AMMI1 models, the test for nonadditivity is $H_0 : d_1 = 0$ vs. $H_a : d_1 \neq 1$, which is equivalent to H_0: $\mathrm{rank}(\boldsymbol{\Gamma}) = 0$ vs. H_a: $\mathrm{rank}(\boldsymbol{\Gamma}) = 1$. Let $\boldsymbol{H}_\nu$ be the row-space or column-space projection operator, $\boldsymbol{H}_\nu = \boldsymbol{I}_\nu - (1/\nu)\mathbf{1}_\nu \mathbf{1}_\nu'$ ($\nu = I, J$), and let $\boldsymbol{K}_\nu \boldsymbol{K}_\nu'$ be a full-rank factorization of $\boldsymbol{H}_\nu$ with dimension $\nu \times (\nu - 1)$ satisfying $\boldsymbol{K}_\nu' \boldsymbol{K}_\nu = \boldsymbol{I}$. We have

$$\boldsymbol{\Gamma} = \boldsymbol{H}_I \boldsymbol{\Gamma} \boldsymbol{H}_J = \boldsymbol{K}_I \boldsymbol{K}_I' \boldsymbol{\Gamma} \boldsymbol{K}_J \boldsymbol{K}_J' = \boldsymbol{K}_I \boldsymbol{\Phi} \boldsymbol{K}_J', \quad \boldsymbol{\Phi} = \boldsymbol{K}_I' \boldsymbol{\Gamma} \boldsymbol{K}_J,$$

with $\mathrm{rank}(\boldsymbol{\Gamma}) = \mathrm{rank}(\boldsymbol{\Phi}) = r \leq p = \min(I - 1, J - 1)$. The elements of $\boldsymbol{\Phi}$ form a basis for the set of interaction contrasts. Define $\boldsymbol{K} = (\boldsymbol{K}_J \otimes \boldsymbol{K}_I)$ so that $\boldsymbol{K}'\mathrm{vec}(\bar{\boldsymbol{Y}})$ is a linear function of $\mathrm{vec}(\bar{\boldsymbol{Y}})$ without containing the main effects (because $\boldsymbol{K}'\mathrm{vec}(\boldsymbol{L}) = 0$). Hence, $\boldsymbol{E}[\boldsymbol{K}'\mathrm{vec}(\bar{\boldsymbol{Y}})] = \boldsymbol{K}'\mathrm{vec}(\boldsymbol{\Gamma}) = \mathrm{vec}(\boldsymbol{\Phi})$ and $\mathrm{Var}(\boldsymbol{K}'\mathrm{vec}(\bar{\boldsymbol{Y}})) = \boldsymbol{K}'\mathrm{Diag}(\delta_{ij}^2)\boldsymbol{K}$.

The goal is to maximize the likelihood function of $\bar{\boldsymbol{Y}}$ subject to the constraint $\mathrm{rank}(\boldsymbol{\Gamma}) = r$, which is the same as computing

$$S(r) = \min_{\mathrm{rank}(\boldsymbol{\Phi}) = r} [\boldsymbol{K}'\mathrm{vec}(\bar{\boldsymbol{Y}}) - \mathrm{vec}(\boldsymbol{\Phi})]' \boldsymbol{W}^{-1} [\boldsymbol{K}'\mathrm{vec}(\bar{\boldsymbol{Y}}) - \mathrm{vec}(\boldsymbol{\Phi})], \tag{4}$$

where $\boldsymbol{W} = \boldsymbol{K}'\mathrm{Diag}(\delta_{ij}^2)\boldsymbol{K}$, and δ_{ij}^2 is replaced by $\hat{\delta}_{ij}^2$ in (3). The constrained ML estimate $\hat{\boldsymbol{\Phi}}$ is the solution to $S(r)$. Due to the weight matrix $\boldsymbol{W}$, a direct SVD solution does not exist. Instead, $\hat{\boldsymbol{\Phi}}$ can be obtained by criss-cross regression [Gabriel and Zamir, 1979]. Write $\boldsymbol{\Phi} = \boldsymbol{A}\boldsymbol{B}'$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are $(I - 1) \times r$ and $(J - 1) \times r$, respectively. Now (4) becomes a standard weighted least squares problem. Given $\boldsymbol{A}^{(n)}$, $\boldsymbol{B}$ is updated as

$$\boldsymbol{B}^{(n+1)} = [(\boldsymbol{I}_{J-1} \otimes \boldsymbol{A}^{(n)})' \boldsymbol{W}^{-1} (\boldsymbol{I}_{J-1} \otimes \boldsymbol{A}^{(n)})]^{-1} \times (\boldsymbol{I}_{J-1} \otimes \boldsymbol{A}^{(n)})' \boldsymbol{W}^{-1} \boldsymbol{K}' \mathrm{vec}(\bar{\boldsymbol{Y}}) \quad (5)$$

In turn, given $\boldsymbol{B}^{(n+1)}$, $\boldsymbol{A}$ is updated as

$$\boldsymbol{A}^{(n+1)} = [(\boldsymbol{B}^{(n+1)} \otimes \boldsymbol{I}_{I-1})' \boldsymbol{W}^{-1} (\boldsymbol{B}^{(n+1)} \otimes \boldsymbol{I}_{I-1})]^{-1} \times (\boldsymbol{B}^{(n+1)} \otimes \boldsymbol{I}_{I-1})' \boldsymbol{W}^{-1} \boldsymbol{K}' \mathrm{vec}(\bar{\boldsymbol{Y}}) \quad (6)$$

We alternate (5) and (6) until convergence of (4) is reached, and $\hat{\boldsymbol{\Phi}} = \hat{\boldsymbol{A}} \hat{\boldsymbol{B}}'$. The LRT statistic is $S(0) - S(1)$, where $S(0) = [\boldsymbol{K}' \mathrm{vec}(\bar{\boldsymbol{Y}})]' \boldsymbol{W}^{-1} \boldsymbol{K}' \mathrm{vec}(\bar{\boldsymbol{Y}})$. The asymptotic null distribution of this LRT statistic converges in distribution to the maximum root of a p-variate Wishart matrix with df = $\max(I-1, J-1)$ in balanced designs [Boik, 1989]. The corresponding 95th and 99th percentiles of this distribution can be found in Hanumara and Thompson Jr [1968]. With unbalanced data, under the assumption that $N_{ij} = \sum_j N_{ij} \sum_i N_{ij} / \sum_{ij} N_{ij}$, the null distribution of the LRT is known to be identical to that in balanced designs. Due to correlated nature of the outcome data, these approximations are not directly applicable to our context. However, our numerical work illustrates that using this reference distribution provides a conservative approximation to the test.

Parameter Estimation Based on Individual Observations

The cell-mean approach provides a quick way of summarizing interaction effects for repeated-measures data. In the presence of confounders and other covariates, a mixed-effects regression model uses all individual observations and provides a general framework for handling repeated measurements. Let $\boldsymbol{y}_{ijk}$ denote the length-n_{ijk} observation vector for subject (i, j, k),

$$\boldsymbol{y}_{ijk} = \mu_{ij} \mathbf{1}_{n_{ijk}} + \boldsymbol{Z}_{ijk} \boldsymbol{b}_{ijk} + \boldsymbol{e}_{ijk}, \quad i = 1, \ldots, I, \quad j = 1, \ldots, J, \quad k = 1, \ldots, N_{ij}, \quad (7)$$

where μ_{ij} is the mean response value for the (i, j)th cell, $\boldsymbol{Z}_{ijk}$ is a $n_{ijk} \times q$ design matrix for the random effects, $\boldsymbol{e}_{ijk} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{ijk})$ are the random errors, not depending on i, j, or k except that its size is $n_{ijk} \times n_{ijk}$, and $\boldsymbol{b}_{ijk}$ is a length-q vector of subject-specific random effects, independent of $\boldsymbol{e}_{ijk}$. The random effects are distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$, where $\boldsymbol{\Psi}$ is the $q \times q$ covariance matrix for the random effects. It follows that the variance-covariance matrix for $\boldsymbol{y}_{ijk}$ is $\boldsymbol{V}_{ijk} = \boldsymbol{Z}_{ijk} \boldsymbol{\Psi} \boldsymbol{Z}'_{ijk} + \boldsymbol{\Sigma}_{ijk}$.

Classical Interaction Models

To avoid computationally intensive iterations associated with ML estimation for models (a)–(d), we propose a two-step regression procedure to approximate the interaction parameters. The idea is similar to Milliken and Johnson [1989], who applied a two-step regression procedure in two-way tables to estimate nonlinear interaction effects. Let $\boldsymbol{X}_{ijk}$ be the design matrix with dimension $n_{ijk} \times IJ$ that allows estimation of all plausible effects from the row and column factors. In the first step, we fit a saturated interaction model to the data using a linear mixed-effects model:

$$\boldsymbol{y}_{ijk} = \boldsymbol{X}_{ijk} \boldsymbol{\xi} + \boldsymbol{Z}_{ijk} \boldsymbol{b}_{ijk} + \boldsymbol{e}_{ijk}, \quad (8)$$

where $\boldsymbol{\xi} = (\mu, R_1, \ldots, R_{I-1}, C_1, \ldots, C_{J-1}, \gamma_{11}, \ldots, \gamma_{(I-1)(J-1)})'$. The log-likelihood function is

$$\ell(\boldsymbol{\xi}, \boldsymbol{\Psi}, \boldsymbol{\Sigma}) = -\frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{N_{ij}} \left[n_{ijk} \log(2\pi) + \log(|\boldsymbol{V}_{ijk}|) + (\boldsymbol{y}_{ijk} - \boldsymbol{X}_{ijk} \boldsymbol{\xi})' \boldsymbol{V}_{ijk}^{-1} (\boldsymbol{y}_{ijk} - \boldsymbol{X}_{ijk} \boldsymbol{\xi}) \right]. \quad (9)$$

The variance components are estimated by restricted maximum likelihood (REML) [Patterson and Thompson, 1971], and the $I \times J$ fixed effect estimates are

$$\hat{\boldsymbol{\xi}} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{N_{ij}} \left(\boldsymbol{X}'_{ijk} \hat{\boldsymbol{V}}_{ijk}^{-1} \boldsymbol{X}_{ijk} \right)^{-1} \boldsymbol{X}'_{ijk} \hat{\boldsymbol{V}}_{ijk}^{-1} \boldsymbol{y}_{ijk}. \quad (10)$$

In the second step, we extract the main effect estimates from $\hat{\boldsymbol{\xi}}$ and compute the residuals

$$\boldsymbol{r}_{ijk} = \boldsymbol{y}_{ijk} - \hat{\mu} \mathbf{1}_{n_{ijk}} - \hat{R}_i \mathbf{1}_{n_{ijk}} - \hat{C}_j \mathbf{1}_{n_{ijk}}. \quad (11)$$

Since the interaction term of Tukey's and Mandel's models involves main effects, we perform a second regression (without intercept) where the residuals $\boldsymbol{r}_{ijk}$ are treated as the response variable and the respective specific forms of main effect estimates are treated as the regressors to obtain the corresponding slope estimates. The second-step regression equations for models (a)–(c) are respectively

$$\boldsymbol{r}_{ijk} = \theta \hat{R}_i \hat{C}_j \mathbf{1}_{n_{ijk}} + \boldsymbol{\epsilon}_{ijk} \quad (12)$$

$$\boldsymbol{r}_{ijk} = \lambda_i \hat{C}_j \mathbf{1}_{n_{ijk}} + \boldsymbol{\epsilon}_{ijk} \quad (13)$$

$$\boldsymbol{r}_{ijk} = \hat{R}_i \eta_j \mathbf{1}_{n_{ijk}} + \boldsymbol{\epsilon}_{ijk} \quad (14)$$

with $\boldsymbol{\epsilon}_{ijk} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\boldsymbol{\epsilon}(n_{ijk} \times n_{ijk})})$. One can select a covariance structure $\boldsymbol{\Omega}_{\boldsymbol{\epsilon}}$ depending on the criterion of model fitting. Note that parameter constraints in (2) are handled in the regressors, and there are $I-1$ and $J-1$ regression equations in (13) and (14), respectively. For model (d), we first obtain the estimates of λ_i and η_j then compute the second-step residuals using $\{\hat{R}_i, \hat{C}_j, \hat{\lambda}_i, \hat{\eta}_j\}$, and finally estimate θ (see Supporting Information for details).

AMMI1 Model

Given that the interaction structure of model (e) is derived from an SVD of the matrix of residuals after removing additive effects, we propose to perform SVD to the saturated $\hat{\boldsymbol{\Gamma}}$ matrix as obtained from $\hat{\boldsymbol{\xi}}$ in (10). The resulting largest singular value of $\hat{\boldsymbol{\Gamma}}$ is an approximation of $\hat{d}_1$. The corresponding left and right singular vectors are approximations of $\hat{\alpha}_i$ and $\hat{\beta}_j$, for $i = 1, \ldots, I$, $j = 1, \ldots, J$.

Remark 1. We evaluated the bias and mean squared error properties of the two-step regression estimators through simulation. The empirical results indicate that the two-step regression estimators appear to be unbiased, even under misspecified correlation structures (Supplementary Table S2). The estimator of d_1 for AMMI1 models (obtained by SVD of

the estimated saturated interaction matrix), however, slightly overestimates d_1.

Parametric Bootstrap Using a LRT Pivot

We construct a LRT statistic based on the noniterative two-step regression estimates. First, the log-likelihood under the null hypothesis is obtained by fitting an additive mixed model (denoted as $\hat{\ell}_0$), and the log-likelihood under the alternative hypothesis is obtained by the previous two-step regression procedure (denoted as $\hat{\ell}_1$). Specifically for calculating $\hat{\ell}_1$, we extract $\hat{\mu}$, $\hat{R}_i$, and $\hat{C}_j$ from (10) and obtain the interaction effect estimates from the second-step regression of residuals on a prespecified structure of main effect estimates for models (a)–(d). Subsequently, an approximate LRT pivot is created, $\tilde{\Lambda} = -2(\hat{\ell}_0 - \hat{\ell}_1)$. Because the parameter estimates used in $\tilde{\Lambda}$ are not proper ML estimates, the resulting test statistic does not have a standard asymptotic distribution. We use parametric bootstrap to elicit the null distribution of this LRT-based pivot. Since permuting Y or subjects across the configurations of G and E factors can remove both interaction and main effects, we generate pseudo data $\boldsymbol{y}^*_{ijk}$ under the null hypothesis of no interaction while preserving the main effects using the model: $\boldsymbol{y}^*_{ijk} = \hat{\mu}\mathbf{1}_{n_{ijk}} + \hat{R}_i\mathbf{1}_{n_{ijk}} + \hat{C}_j\mathbf{1}_{n_{ijk}} + \boldsymbol{Z}_{ijk}\boldsymbol{b}^*_{ijk} + \boldsymbol{e}^*_{ijk}$, where $\boldsymbol{b}^*_{ijk} \sim N(0, \hat{\boldsymbol{\Psi}})$, $\boldsymbol{e}^*_{ijk} \sim N(0, \hat{\boldsymbol{\Sigma}})$. $\hat{\boldsymbol{\Psi}}$ and $\hat{\boldsymbol{\Sigma}}$ are REML estimates from the saturated mixed-effects model in (8). For each simulated null sample, a $\tilde{\Lambda}$ is computed. Repeating the procedure for a large number of times (e.g., we use 1,000) provides an approximate distribution of $\tilde{\Lambda}$ under H_0. Finally, an empirical P-value is obtained by calculating the proportion of all $\tilde{\Lambda}$ that exceeds the observed $\tilde{\Lambda}$.

Simulation Settings

We carried out a series of simulation studies to examine the following properties of the proposed tests: (1) type I error for the LRT using cell means (LRT-CM) and for the parametric bootstrap approach with the LRT-based statistic $\tilde{\Lambda}$ (LRT-PB); (2) power comparison of AMMI1 to saturated interaction model if AMMI1 model holds; (3) power comparison of LRT-PB to LRT-CM; and (4) performance of models (a)–(e) across classical interaction structures. In addition, we compared relative performances of our proposed tests to existing strategies for testing GGI or GEI, including the standard saturated interaction model and the naive cell-mean approach of Barhdadi and Dubé [2010] (not accounting for correlated data). Furthermore, we evaluated the performance of each model in detecting GGI with repeated measures on a quantitative trait under 12 epistasis patterns.

Individual-level outcome Y with n_{ijk} repeated measures on subject (i, j, k) were generated for a total of N subjects. The general description of the model is given by

$$\boldsymbol{y}_{ijk} = \mu_{ij}\mathbf{1}_{n_{ijk}} + b_{ijk}\mathbf{1}_{n_{ijk}} + \boldsymbol{e}_{ijk}, \quad k = 1, \ldots, N_{ij}, \qquad (15)$$

where $\boldsymbol{e}_{ijk} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2\boldsymbol{I})$, $b_{ijk} \sim \mathcal{N}(0, \sigma_b^2)$, and $\{\boldsymbol{e}, \boldsymbol{b}\}$ are mutually independent. Cell means were first generated according to models (a)–(e), and the data vector for each individual with a given mean and a covariance structure was generated from a multivariate normal distribution. The interaction terms in all models were scaled in such a way that they contributed to 15% of the total variation explained by the model, and the remainder was attributed to row and column main effects. While simulating data under an AMMI1 model, we assigned the entire contribution due to interaction effect to the first interaction factor.

To evaluate model performance in terms of detecting common patterns of GGI, data were simulated according to 12 epistasis models [Barhdadi and Dubé, 2010]: (1) dominant or dominant; (2) dominant or recessive; (3) modified; (4) dominant and dominant; (5) recessive or recessive; (6) threshold; (7) dominant and recessive; (8) recessive and recessive ((1)–(8) from Jung et al. [2009]); (9) checkboard; (10) additive and additive; (11) diagonal; ((9)–(11) from Culverhouse et al. [2004]); and (12) general. The additive and additive model and the general model are purely epistatic models, that is, the quantitative trait depends on genotype from two loci in the absence of any marginal effects. Figure 1 gives a visual representation of the true cell means for the 12 epistasis patterns in our simulation.

We considered 3 × 3 table settings for all simulations to mimic studies of GGI. In addition, we evaluated the power of AMMI1 models in 9 × 5 table settings as described in [2]. For 3 × 3 tables, minor allele frequencies for the two loci were set at 0.3 and 0.4, respectively. For 9 × 5 tables, combinations of the two loci with allele frequencies 0.3 and 0.4 (resulting in nine categories) along with an environmental exposure with five levels (each with probability 0.2) were considered.

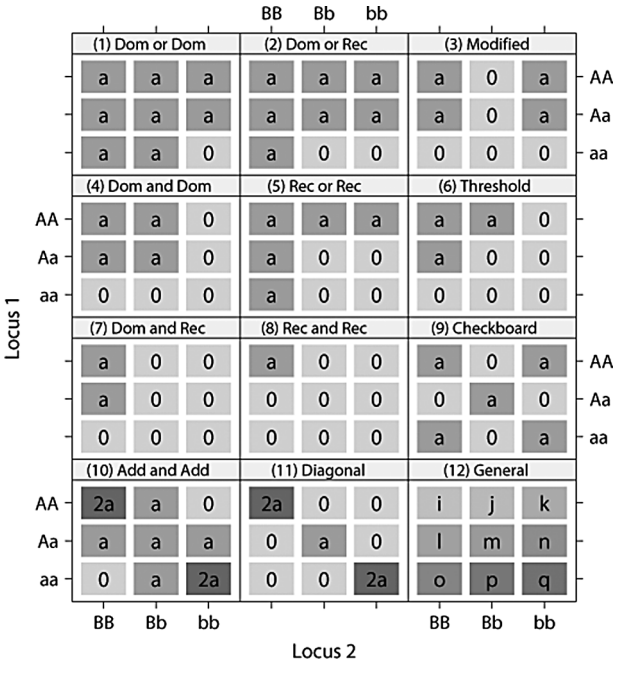

Figure 1. Cell means (a = 0.5) for 12 common epistasis models in the simulation setting.

Hardy-Weinberg equilibrium was assumed to hold for both loci. We set (i) $\sigma^2 = 8$, $\rho = 0.5$ (or $\sigma_b^2 = \sigma_e^2 = 4$) and (ii) $\sigma^2 = 16$, $\rho = 0.5$ (or $\sigma_b^2 = \sigma_e^2 = 8$). We also considered $\rho = \{0.2, 0.5, 0.8\}$ for the power evaluation of AMMI1 models. Under each simulation setting, 1,000 datasets were generated with 1,800 and 3,600 subjects for 3×3 and 9×5 tables, respectively. The number of repeated measurements per subject was generated from a multinomial distribution similar to the analysis dataset: $n_{ijk} \in \{2, 3, 4, 5, 6\}$, $\boldsymbol{n} = \{n_{ijk} : 1 \le k \le N_{ij}, 1 \le i \le I, 1 \le j \le J\} \sim \text{mult}(N, \boldsymbol{p})$, $\boldsymbol{p} = (0.15, 0.2, 0.3, 0.2, 0.15)$. This is equivalent to generating outcome data missing completely at random.

Results

Simulation Findings

Type I Error

We generated data under an additive model, $H_0 : \gamma_{ij} = 0$ (while $R_i, C_j \neq 0$) as well as under a completely null model, $H_0 : \gamma_{ij} = R_i = C_j = 0$ for all i, j. Figure 2 shows the percentage of false rejections for the five interaction models from 1,000 simulations at 5% significance level. Under the additive model, the type I error rates for all models using LRT-CM and LRT-PB are maintained at the nominal 5%. Under the null model, type I error rates for models (a)–(e) using LRT-PB as well as for model (e) using LRT-CM are still maintained at 5%. LRT-CM for classical models (a)–(d), however, are either too liberal or too conservative (>12% for model (a), >8% for models (b) and (c), and <3% for model (d)).

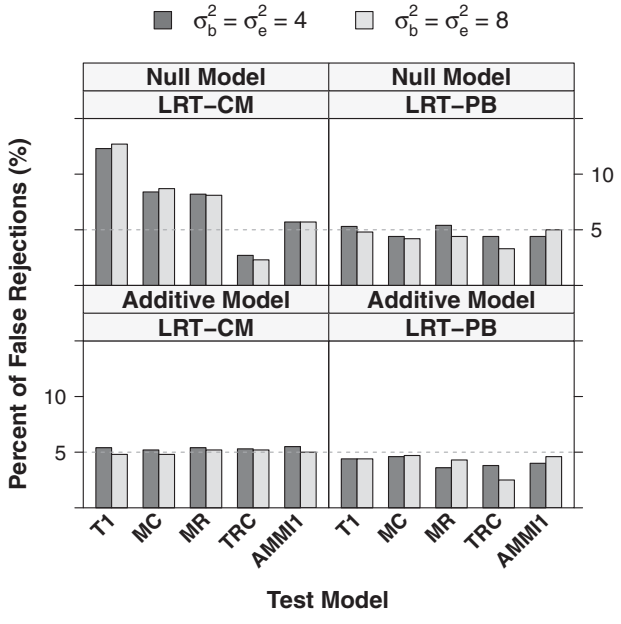

Figure 2. Type I error for the five interaction tests in a 3×3 array setting using the likelihood ratio test with the cell-mean approach (LRT-CM) and the parametric bootstrap test (LRT-PB). A total of 1,000 simulation datasets are generated under an additive model (only main effects) and under a completely null model (no main or interaction effects). T1 = Tukey's one degree-of-freedom nonadditivity test model (a), MC = Mandel's column model (b), MR = Mandel's row model (c), TRC = Tukey's row-column model (d), AMMI1 = model (e).

Power

The gain in power using an AMMI1 model compared to a saturated interaction model increases as the table dimension increases. In the 3×3 array setting, saturated models (4 df for the interaction effects) appear to have similar power to AMMI1 models (data not shown). In the 9×5 array setting, AMMI1 models using LRT-CM and LRT-PB clearly have greater power than saturated models when the true interaction only has one interaction factor (Fig. 3). The highest observed gain in power for AMMI1 using LRT-PB compared to the saturated model (32 df for the interaction effects) is 11% under three correlation settings. As ρ increases from 0.2 to 0.8, AMMI1 begins to show power gain across a wider range of d_1.

Figure 4 shows the percentage of interactions detected by each test across a set of true simulation models. Overall, the power of LRT-PB is increased by 2–5% compared to LRT-CM. When Tukey's model (a) is the true model, all other models are able to capture some interactions (70–82% when $\sigma_b^2 = \sigma_e^2 = 4$, 38–53% when $\sigma_b^2 = \sigma_e^2 = 8$). Under Mandel's column model (b), Tukey's row-column (d) and AMMI1 (e) are able to detect the interaction (both power >99% when $\sigma_b^2 = \sigma_e^2 = 4$ and >90% when $\sigma_b^2 = \sigma_e^2 = 8$); whereas Tukey's 1-df model (a) and Mandel's row model (c) have very low power (both <50% with LRT-CM and <6% with LRT-PB). Similar properties are observed for simulations under Mandel's row model (c). With Tukey's row-column model (d) being the simulation model, all alternatives, except model (a), are able to detect the interaction with power greater than 60%. When the true model is an AMMI1 model (e), models (a)–(c) have relatively low power to detect interaction (<50% when $\sigma_b^2 = \sigma_e^2 = 4$ and <32% when $\sigma_b^2 = \sigma_e^2 = 8$). Saturated model has lower power than AMMI1 in most cases.

Figure 5 shows the percentages of interaction detected by six interaction models using LRT-PB under 12 common epistasis models. Given the robust performance of model (e), AMMI1 model appears to be a desirable approach for evaluating common epistasis structures, especially when main effects do not exist (e.g., epistasis models (10) and (12)).

We also compared our proposed methods with those in Barhdadi and Dubé [2010] in terms of type I error and power. As expected, the tests in Barhdadi and Dubé [2010] assuming balanced data structure and not accounting for within-subject correlation yield inflated type I error (especially for Tukey's and Mandel's models) and lower power (see Supplementary Fig. S2).

Application to the Normative Aging Study (NAS)

The NAS is a multidisciplinary longitudinal study initiated by the U.S. Veterans Administration in 1963 [Bell et al.,

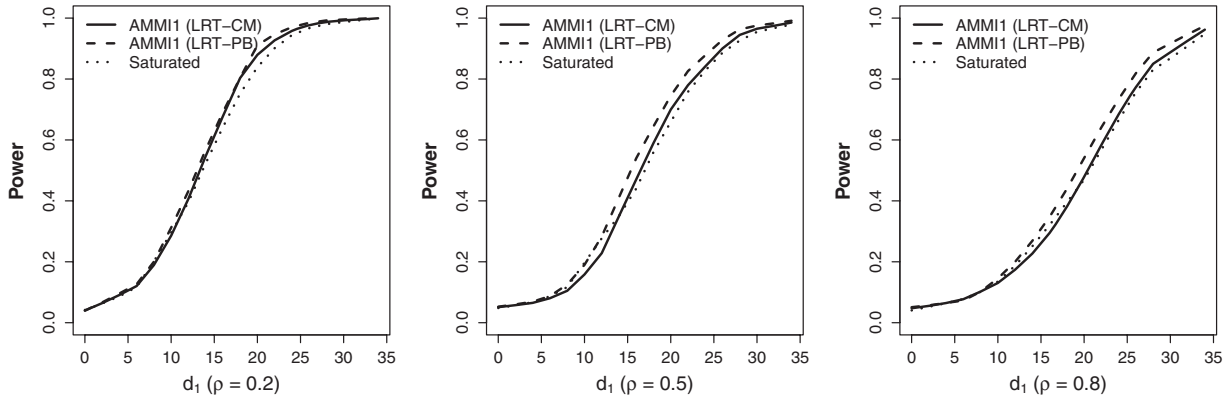

Figure 3. Empirical power (or true positive rate) of AMMI1 model (at $\alpha = 0.05$) using the likelihood ratio test with the cell-mean approach (LRT-CM) and the parametric bootstrap test (LRT-PB), and a saturated interaction model in a 9×5 array setting with $\sigma^2 = 8$ and $\rho = 0.2$, 0.5, and 0.8.

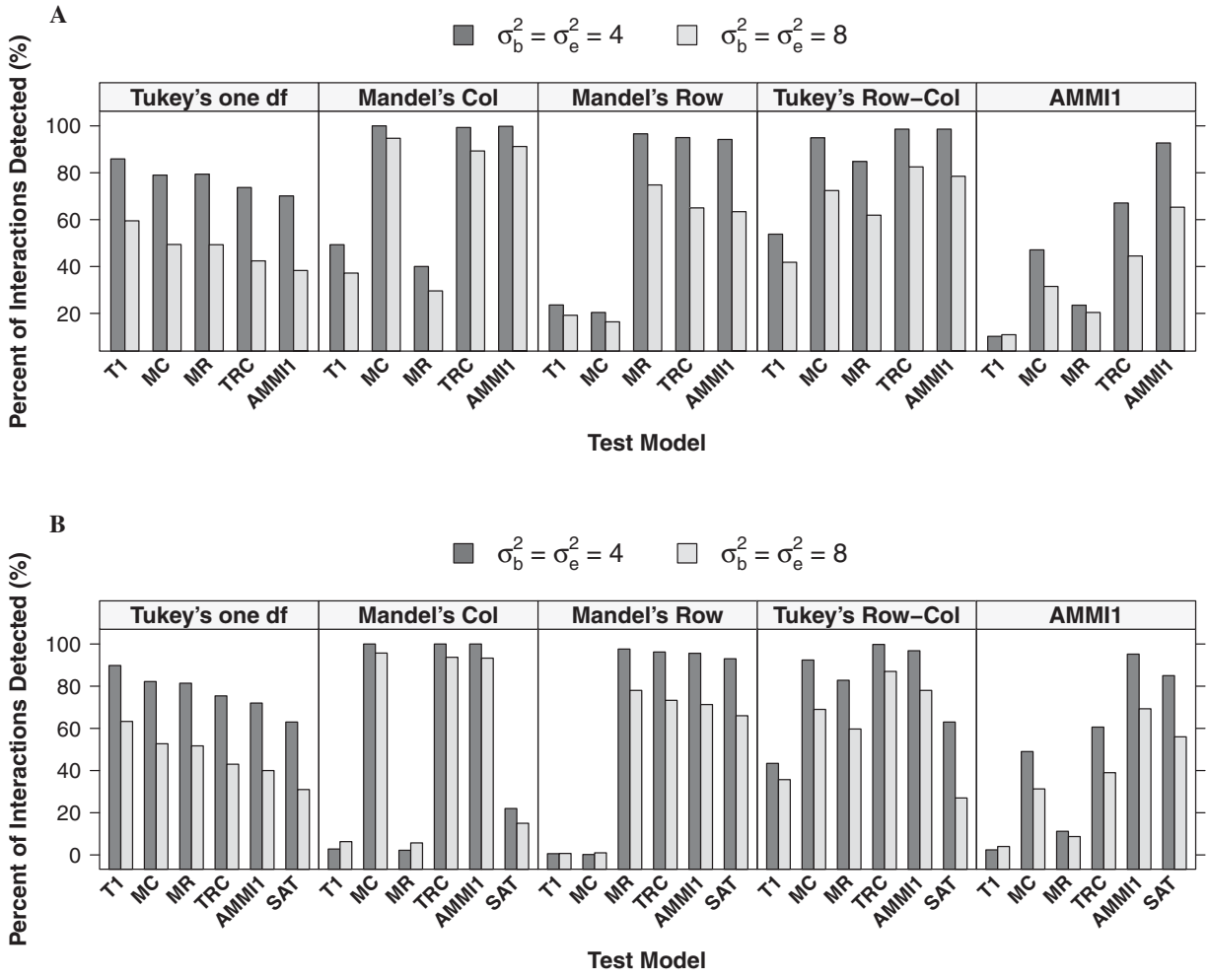

Figure 4. Percentage of interactions detected by different interaction models in the simulation settings corresponding to a 3×3 array. Results are based on (A) the likelihood ratio test with the cell-mean approach (LRT-CM) and (B) the parametric bootstrap test (LRT-PB) with test results of using a saturated model for interaction as a comparison. The top label within each box represents the true simulation model. The horizontal-axis labels indicate the models used for testing interaction. T1 = Tukey's one degree-of-freedom nonadditivity test model (a), MC = Mandel's column model (b), MR = Mandel's row model (c), TRC = Tukey's row-column model (d), AMMI1 = model (e), SAT = saturated interaction model.

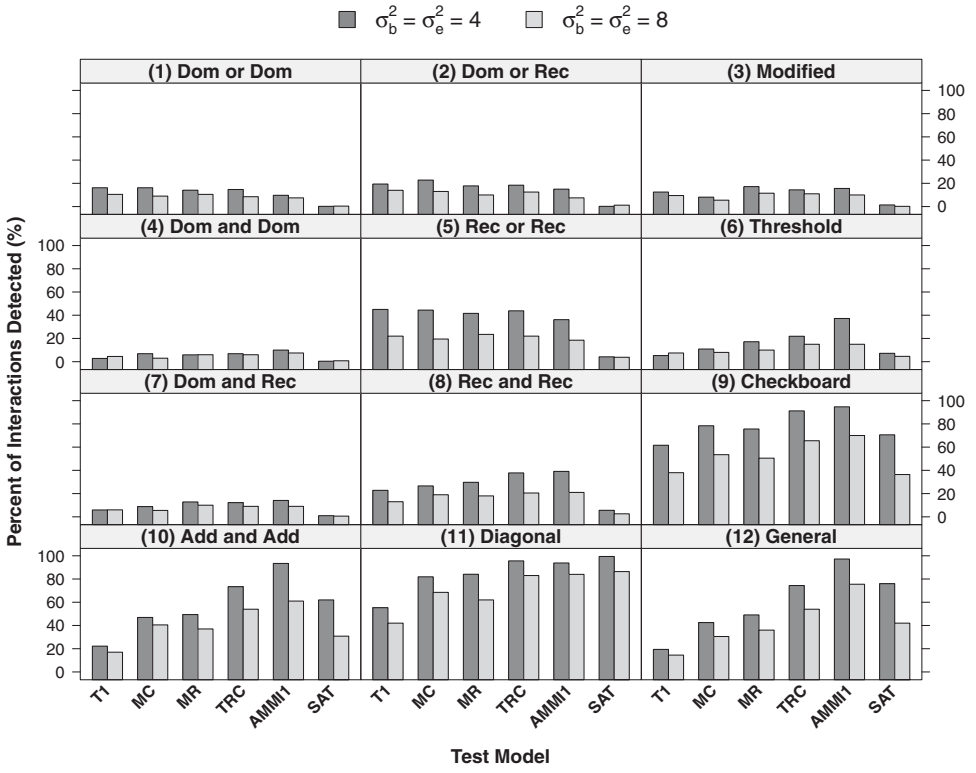

Figure 5. Percentage of interactions detected (or null hypotheses of no interaction rejected) by each of the interaction models using parametric bootstrap test (LRT-PB) and a saturated model for interaction under 12 common epistasis models. T1 = Tukey's one degree-of-freedom nonadditivity test model (a), MC = Mandel's column model (b), MR = Mandel's row model (c), TRC = Tukey's row-column model (d), AMMI1 = model (e), SAT = saturated interaction model.

1966]. We analyzed 671 participants from a subset of the NAS data who were successfully genotyped for the *HFE* gene and had baseline measurements of tibia bone lead (a measure of cumulative lead exposure) [Zhang et al., 2010]. The analysis goal was to investigate effect modification by the different *HFE* alleles on the association between lead exposure and PP, which is a strong predictor of heart problems for older adults. Since 1991, data had been collected every 3–5 years until 2011 with a median follow-up time of 12 years, including physical examination, blood pressure and laboratory measurements, and questionnaire data. The majority (97%) of the participants were Caucasian. The average age was 66.29 ± 7.14 (range 48–93) at the time of tibia bone lead measurement. More than 96% of subjects had repeated measurements on blood pressure, and over 65% of them had at least four measurements during the study period contributing to a total of 2,914 observations.

Two major mutations in the *HFE* gene (*C282Y* and *H63D* mutations) were considered for analysis following Zhang et al. [2010]. Let (*AA, Aa, aa*) and (*BB, Bb, bb*) denote wild type, having one variant allele, and having two variant alleles for *C282Y* and *H63D*, respectively. As a result of small sample sizes in certain homozygote genotypes ($N = 5$ for *aaBB*, $N = 17$ for *AAbb*) and compound heterozygotes ($N = 14$ for *AaBb* and $N = 0$ for *Aabb, aaBb, aabb*), we were unable to test GGI between the two loci. Since the research interest was to compare three mutually exclusive groups (wild type, *H63D, C282Y*), 14 subjects with compound heterozygotes (*AaBb*) were excluded from analysis. Consequently, the *HFE* genotypes were classified into three categories for analysis: *AABB, AaBB* or *aaBB*, and *AABb* or *AAbb*. The environmental exposure (cumulative lead) was a continuous variable, but to illustrate the proposed methods, we categorized bone lead levels into three groups (low: ≤15, medium: (15, 25], and high: >25 $\mu g/g$). Table 1 lists the observed cell means of PP and the number of participants for each $G \times E$ configuration.

Table 1. Cell means corresponding to pulse pressure and number of participants (in parentheses) for each configuration of the *HFE* genotypes and bone lead levels in the Normative Aging Study

	Tibia lead levels (μg/g)		
HFE gene	Low: ≤15	Medium: >15 and ≤25	High: >25
Wild-type *(AABB)*	52.94 (161)	56.16 (149)	56.61 (131)
C282Y (AaBB or *aaBB)*	51.89 (23)	56.65 (39)	59.10 (23)
H63D (AABb or *AAbb)*	52.58 (54)	57.72 (53)	64.49 (38)

We applied LRT-CM and LRT-PB to test this GEI effect. According to the Akaike information criterion for model fit, we chose a random-intercept mixed-effects model for analysis:

Table 2. Analysis results of GEI between *HFE* genotypes and tibia lead levels in the Normative Aging Study (*N* = 671) using the proposed likelihood ratio test with cell means (LRT-CM) and the parametric bootstrap (LRT-PB) approach (1,000 replicates simulated under the null hypothesis)

		P-value			
Model	Hypothesis	LRT-CM[a]	LRT-CM[b]	LRT-PB[a]	LRT-PB[b]
Model (a)	$H_0 : \theta = 0$	0.008	0.002	0.002	0.003
Model (b)	$H_0 : \lambda_i = 0$ (lead)	0.029	0.007	0.009	0.008
Model (c)	$H_0 : \eta_j = 0$ (*HFE*)	0.015	0.002	0.002	0.001
Model (d)	$H_0 : \theta = \lambda_i = \eta_j = 0$	0.035	0.005	0.007	0.002
Model (e)	$H_0 : d_1 = 0$	<0.10	<0.01	0.009	0.002
Saturated model	*HFE* × lead			0.015	0.006

[a] No covariate adjustment.
[b] Adjusting for baseline age, time since baseline, and squared time. For LRT-CM, residuals from a regression of pulse pressure on all other covariates except lead levels and genotype were used to form the cell means.

$$\boldsymbol{y}_{ijk} = \mu_{ij}\mathbf{1}_{n_{ijk}} + b_{ijk}\mathbf{1}_{n_{ijk}} + \boldsymbol{e}_{ijk}, \quad \text{with} \quad \mu_{ij} = \mu + R_i + C_j + \gamma_{ij}, \tag{16}$$

where $b_{ijk} \sim \mathcal{N}(0, \sigma_b^2)$ is the random-effect coefficient for subject (i, j, k), $\boldsymbol{e}_{ijk} \sim \mathcal{N}(\mathbf{0}, \sigma_e^2\boldsymbol{I})$ is the random error term, and $\{\sigma_b^2, \sigma_e^2\}$ are assumed to be constant across individuals. We also considered the model adjusting for baseline age, time since baseline in years, and squared time.

$$\boldsymbol{y}_{ijk} = \mu_{ij}\mathbf{1}_{n_{ijk}} + \beta_1 \text{Age}_{ijk} + \beta_2 \mathbf{Time}_{ijk} + \beta_3 \mathbf{Time}^2_{ijk} + b_{ijk}\mathbf{1}_{n_{ijk}} + \boldsymbol{e}_{ijk}. \tag{17}$$

For LRT-CM adjusting for covariates, cell means were formed by the residuals from a regression of the outcome on covariates other than *G* and *E*. This is an ad hoc approach for covariate adjustment since correlations of covariates with *G* and *E* are ignored. In general, LRT-PB based on a full regression model with *G*, *E*, and covariates will yield more power.

Results for GEI

Table 2 shows the *P*-values for testing *HFE* × lead exposure interaction using LRT-CM and LRT-PB and the saturated interaction model. Using LRT-CM without adjustment for any covariate, the interaction was significant in all four classical models ($P < 0.05$), whereas model (e) gave a *P*-value between 0.05 and 0.10. After adjusting for the covariates, model (e) detected the interaction using LRT-CM ($P < 0.01$). Regardless of covariate adjustment, the interaction was significant for models (a)–(e) using LRT-PB ($P < 0.01$), and also for the saturated interaction model ($P < 0.02$). *P*-values for the GEI effect decreased further for all tests with adjustment for baseline age, time since baseline, and squared time. Given the significant GEI on all models, this interaction appears to be present irrespective of model choice.

According to the SVD of $\hat{\boldsymbol{\Gamma}}_{G\times E}$ under a saturated interaction model with covariate adjustment, the first and second characteristic roots of $\hat{\boldsymbol{\Gamma}}_{G\times E}$ were $\hat{d}_1 = 5.65$ and $\hat{d}_2 = 1.24$, respectively (Supplementary Table S3). The first interaction factor contributed to $\hat{d}_1^2/(\hat{d}_1^2 + \hat{d}_2^2) = 95\%$ of the total contribution to the interaction term. The association between PP and bone lead levels was strongest among *H63D* variant (*AABb* or *AAbb*) carriers, compared to *C282Y* variant (*AaBB* or *aaBB*) carriers and wild-type (*AABB*) participants. Based on the saturated interaction model estimates, the estimated difference in mean PP for *H63D* variant carriers with high vs. low lead levels was 9.6 mmHg [95% confidence interval (CI), 0.43–14.83 mmHg]. The same estimated differences were 3.52 mmHg [95% CI, –3.39–10.43 mmHg] and –0.33 mmHg [95% CI, –2.95–2.29 mmHg] for *C282* variant carriers and wild-type participants, respectively. Although the underlying mechanisms are poorly understood, the common prooxidant and proinflammatory effects shared by lead exposure and *H63D HFE*, especially through increased production of monocyte chemoattractant protein-1 [Zhang et al., 2010], may explain the observed *H63D HFE*-lead interaction.

Subject-Specific and Age-Specific Contributions to GEI

Using the estimates of singular vectors $(\hat{\alpha}_{im}, \hat{\beta}_{jm})$, we investigated subject-specific and age-specific contributions to GEI in the first and the second interaction factors ($m = 1, 2$) via the sum of squared deviations [Mukherjee et al., 2012]. Briefly, the variation due to subject (i, j, k) can be calculated by $\hat{d}_{ijkm} = \hat{\alpha}_{im}\hat{\beta}_{jm}\hat{r}_{ijk\cdot}$, where $\hat{r}_{ijk\cdot}$ is the mean of n_{ijk} subject-level residuals from (11) after removing main effects. The variation in the contribution of subject (i, j, k) is then $(\hat{d}_{ijkm} - \hat{d}_{\cdot m})^2$, where $\hat{d}_{\cdot m} = \sum_i \sum_j \sum_k \hat{d}_{ijkm}/N$. For the age-specific contribution, we constructed eight 3-year age intervals. The first and last intervals contained observations from those who were younger than 65 and who were 83 or older, respectively. The cell means of PP and numbers of participants for genotypes and lead levels based on different age categories are presented in Supplementary Figure S3. The contribution due to the *t*th age interval is calculated as $\hat{d}_{tm} = \sum_i \sum_j \hat{\alpha}_{im}\hat{\beta}_{jm}\hat{r}_{tij\cdot}$, where $\hat{r}_{tij\cdot}$ is the average of residuals (11) in the *t*th age interval among individuals in the (i, j)th cell, $t = 1, \ldots, 8$. The variation in the contribution of the *t*th interval is $(\hat{d}_{tm} - \hat{d}_{\cdot m})^2$, where $\hat{d}_{\cdot m} = \sum_{t=1}^{8} \hat{d}_{tm}/8$.

Figure 6 displays subject-specific contributions from the 671 individuals (left panel) and contributions of eight age intervals to the first interaction factor of GEI (right panel). The plot indicates that the modifying effect of the *HFE* gene on the effect of cumulative lead exposure on PP spiked around age 75. This was due to the fact that the mean difference in PP between the low and the high bone lead groups became largest in that age interval with *H63D* (*AABb* or *AAbb*), whereas the difference in PP among those with wild type of *HFE* (*AABB*) was the smallest. A stratified analysis by baseline age also indicates time-dependent evidence of interaction effects (see Supporting Information). Supplementary Figure S4 shows patterns corresponding to the second interaction factor with substantially less subject-specific and age-specific variability for the GEI. Although these graphical diagnostics do not establish a statistical significance of a $G \times E \times$ *Time* term,

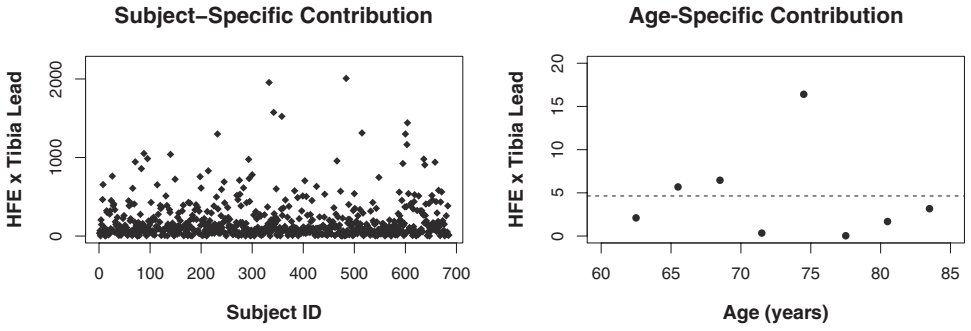

Figure 6. Subject-specific contributions (left) and age-specific contributions (right) to the *first* interaction factor in the *HFE* × lead interaction based on the Normative Aging Study data.

they provide important insight into longitudinal features of the interaction factors.

Discussion

We have proposed new LRTs for GGI and GEI effects in longitudinal cohort studies using a sparse representation of interaction structure via Tukey's and Mandel's models as well as AMMI1 models. AMMI1 appears to be a robust and flexible model in detecting interaction effects across a spectrum of interaction structures. Moreover, it is relatively powerful in detecting certain epistasis structures with no appreciable main effects but potential interaction. In contrast, Tukey's and Mandel's models fail to detect interactions if the interaction structure is misspecified.

Both of our approaches require prior assumptions of the mean structure under the null hypothesis of no interaction and an underlying correlation structure for within-subject measurements. When either part of the model is misspecified, the power and the false rejection rate might be affected. Although this is a generic limitation of parametric modeling, we performed additional simulations to evaluate the influence of misspecification of covariance structure on the proposed tests. We generated data under several common correlation structures (e.g., compound symmetry, autoregressive, unstructured) and analyzed interactions assuming a different correlation structure. The results show that under a misspecification of covariance structure, type I error rates are maintained for LRT-PB but can be slightly inflated for LRT-CM.

In our simulation studies, we did not see a vast difference in the power between LRT-CM and LRT-PB. The correlation across repeated measurements in the LRT-CM approach is accounted for by the weight matrix $\boldsymbol{W}$. Therefore, the test is not based on naive subject-level averages as in Mukherjee et al. [2012]. The main advantage of LRT-PB is the flexible regression structure that allows all readily available mixed model estimation tools to be used.

We have focused on developing valid tests for the five interaction models, yet some limitations must be acknowledged. First, a caveat of the ML estimation for classical interaction models (a)–(d) based on cell means is that when the underlying main effects are relatively small, the estimation for interaction parameters would become numerically unstable. Depending on initial values, the final converged estimates might be local ML estimates instead of global ML estimates. Second, SVD of the estimated saturated interaction matrix yields approximate estimates rather than proper ML estimates for AMMI model parameters. Our simulation results indicate that this estimator leads to slight overestimation of d_1. Nevertheless, LRT-PB for AMMI1 still maintains the nominal type I error rate and in general possesses greater power than a saturated interaction model. In addition, how to connect the parameters of the AMMI model to directly interpretable quantities for biological interactions is not clear. At this stage, AMMI model remains a screening strategy for testing nonadditivity. Third, covariate adjustment was not considered in our simulation studies. In practice, one can incorporate time effect and other (time-varying) covariates with LRT-CM and LRT-PB, as we did in our data example. Lastly, to our knowledge, no replication study has examined the interaction effect between the *HFE* gene and cumulative lead exposure on PP. We randomly split the data in half and analyzed the two halves for GEI as an assessment of internal consistency, and the results were consistent with our findings. As discussed in Zhang et al. [2010], the conclusion from the NAS data analysis may not be generalizable to other populations given that the study population was exclusively white men. Besides, unmeasured confounding factors and interactions with other genetic polymorphisms or environmental factors were not considered.

The proposed analysis strategies are useful for detecting GGI and GEI effects in longitudinal data. A full likelihood-based approach using a general nonlinear mixed-effects model set-up would be more appealing if the appropriate test statistics and their closed form null distributions can be obtained. Further work is required to investigate specialized nonlinear optimization algorithms in the ML framework to replace the two-step estimation and to construct a valid and more efficient test. It is also important to develop a formal test for individual- and time-specific contributions to

interactions, which will ultimately lead to better understanding of GGI and GEI.

Acknowledgments

The research of B.M. was supported by the NSF grant DMS-1007494 and the NIH grants CA156608 and ES020811. S.K.P. was supported by the NIEHS K01-ES016587. The VA Normative Aging Study is supported by the Cooperative Studies Program/Epidemiology Research and Information Center of the U.S. Department of Veterans Affairs and is a component of the Massachusetts Veterans Epidemiology Research and Information Center, Boston, Massachusetts. The authors would like to thank Dr. Joel Schwartz, Dr. Howard Hu, and NAS participants for sharing the data resources. The authors have no conflict of interest to declare.

References

Barhdadi A, Dubé M. 2010. Testing for gene-gene interaction with AMMI models. *Stat Appl Genet Mol Biol* 9:1–27.

Bell B, Rose C, Damon A. 1966. The veterans administration longitudinal study of healthy aging. *Gerontologist* 6:179–184.

Boik R. 1989. Reduced-rank models for interaction in unequally replicated two-way classifications. *J Multivar Anal* 28:69–87.

Bookman E, McAllister K, Gillanders E, Wanke K, Balshaw D, Rutter J, Reedy J, Shaughnessy D, Agurs-Collins T, Paltoo D and others,. 2011. Gene-environment interplay in common complex diseases: forging an integrative model—recommendations from an NIH workshop. *Genet Epidemiol* 35:217–225.

Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S. 2006. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *Am J Hum Genet* 79:1002–1016.

Crossa J, Gauch H, Zobel R. 1990. Additive main effects and multiplicative interaction analysis of two international maize cultivar trials. *Crop Science* 30:493–500.

Culverhouse R, Klein T, Shannon W. 2004. Detecting epistatic interactions contributing to quantitative traits. *Genet Epidemiol* 27:141–152.

Fan R, Albert P, Schisterman E. 2012. A discussion of gene–gene and gene–environment interactions and longitudinal genetic analysis of complex traits. *Stat Med* 31:2565–2568.

Freeman G. 1973. Statistical methods for the analysis of genotype-environment interactions. *Heredity* 31:339–354.

Gabriel K, Zamir S. 1979. Lower rank approximation of matrices by least squares with any choice of weights. *Technometrics* 21:489–498.

Gollob H. 1968. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* 33:73–115.

Hanumara R, Thompson W Jr. 1968. Percentage points of the extreme roots of a Wishart matrix. *Biometrika* 55:505–512.

Jung J, Sun B, Kwon D, Koller D, Foroud T. 2009. Allelic-based gene-gene interaction associated with quantitative traits. *Genet Epidemiol* 33:332–343.

Liang K, Zeger S. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22.

Lin X, Zhang D. 1999. Inference in generalized additive mixed models by using smoothing splines. *J R Stat Soc Ser B Stat Methodol* 61:381–400.

Maity A, Carroll R, Mammen E, Chatterjee N. 2009. Testing in semiparametric models with interaction, with applications to gene–environment interactions. *J R Stat Soc Ser B Stat Methodol* 71:75–96.

Mandel J. 1961. Non-additivity in two-way analysis of variance. *J Am Stat Assoc* 56:878–888.

Mandel J. 1971. A new analysis of variance model for non-additive data. *Technometrics* 13:1–18.

Milliken G, Johnson D. 1989. *Analysis of Messy Data, Volume II: Nonreplicated Experiments*. New York: Van Nostrand Reinhold.

Mukherjee B, Ko Y, VanderWeele T, Roy A, Park S, Chen J. 2012. Principal interactions analysis for repeated measures data: application to gene–gene and gene–environment interactions. *Stat Med* 31:2531–2551.

Nocedal J, Wright S. 1999. *Numerical Optimization*. New York: Springer Verlag.

Patterson H, Thompson R. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554.

R Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Tukey J. 1949. One degree of freedom for non-additivity. *Biometrics* 5:232–242.

Tukey J. 1962. The future of data analysis. *Ann Math Stat* 33:1–67.

Wong M, Day N, Luan J, Chan K, Wareham N. 2003. The detection of gene–environment interaction for continuous traits: should we deal with measurement error by bigger studies or better measurement? *Int J Epidemiol* 32:51–57.

Zhang A, Park S, Wright R, Weisskopf M, Mukherjee B, Nie H, Sparrow D, Hu H. 2010. *HFE H63D* polymorphism as a modifier of the effect of cumulative lead exposure on pulse pressure: the Normative Aging Study. *Environ Health Perspect* 118:1261–1266.

Zobel R, Wright M, Gauch H. 1988. Statistical analysis of a yield trial. *Agron J* 80:388–393.