Principal interactions analysis for repeated measures data: application to gene–gene and gene–environment interactions

Bhramar Mukherjee,[a]*[†] Yi-An Ko,[a] Tyler VanderWeele,[b] Anindya Roy,[c] Sung Kyun Park[d] and Jinbo Chen[e]

Many existing cohorts with longitudinal data on environmental exposures, occupational history, lifestyle/ behavioral characteristics, and health outcomes have collected genetic data in recent years. In this paper, we consider the problem of modeling gene–gene and gene–environment interactions with repeated measures data on a quantitative trait. We review possibilities of using classical models proposed by Tukey (1949) and Mandel (1961) using the cell means of a two-way classification array for such data. Although these models are effective for detecting interactions in the presence of main effects, they fail miserably if the interaction structure is misspecified. We explore a more robust class of interaction models that are based on a singular value decomposition of the cell-means residual matrix after fitting the additive main effect terms. This class of additive main effects and multiplicative interaction models (Gollob, 1968) provide useful summaries for subject-specific and time-varying effects as represented in terms of their contribution to the leading eigenvalues of the interaction matrix. It also makes the interaction structure more amenable to geometric representation. We call this analysis 'principal interactions analysis'. While the paper primarily focuses on a cell-mean-based analysis of repeated measures outcome, we also introduce resampling-based methods that appropriately recognize the unbalanced and longitudinal nature of the data instead of reducing the response to cell means. We illustrate the proposed methods by using data from the Normative Aging Study, a longitudinal cohort study of Boston area veterans since 1963. We carry out simulation studies under an array of classical interaction models and common epistasis models to illustrate the properties of the principal interactions analysis procedure in comparison with the classical alternatives. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: biplot; column interaction; eigenvalue; epistasis; intraclass correlation; likelihood-ratio test; non-additivity; permutation tests; pseudo F-test; row interaction; singular vector; Wishart matrix

1. Introduction

Statistical methods for analysis of interactions are receiving considerable attention in the post-genome-wide association study era where different consortia are examining gene–gene ($G \times G$) and gene–environment ($G \times E$) interactions [1]. Whereas much of the more recent literature has evolved around case–control association studies, less attention has been devoted to longitudinal cohort studies with rich lifetime exposure data and repeated measures on outcomes. Typically, a naive analysis of repeated measures data attempts to model $G \times E$ effects by fitting a regression model to the conditional mean structure of the outcome Y with main effects of G, E, and $G \times E$ terms after adjusting for other confounders. A random intercept term capturing within subject correlation will commonly be introduced

[a] *Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[b] *Department of Epidemiology and Biostatistics, Harvard University, Boston, MA 02115, U.S.A.*
[c] *Department of Mathematics and Statistics, UMBC, Baltimore, MD 21250, U.S.A.*
[d] *Department of Epidemiology, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[e] *Department of Biostatistics, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.*
*Correspondence to: Bhramar Mukherjee, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, U.S.A.
†E-mail: bhramar@umich.edu

in a standard linear mixed model analysis [2, 3] . However, while incorporating longitudinal effects of time in the model for mean response, one is often confronted with the issue of time-varying effects of interaction with a three-way $G \times E \times$ time term turning out to be statistically significant in a routine mixed model analysis. It is hard to interpret the interaction parameter in such instances. One can try to model the time-varying coefficient corresponding to the interaction term in the generalized additive mixed model framework [4], but tests for such nonparametric, smoothed interaction terms will have little or no power for studies with moderate sample size.

In this paper, we first present an alternate approach to explore interaction structures for cohort studies by first considering the average of repeated measures across subjects as a single observation per subject and then examining the cell-mean structure corresponding to the $G = g, E = e$ in a two-way genotype $\times$ environment classification array ($G_1 = g_1, G_2 = g_2$ for a two-way gene $\times$ gene array). Because of the two-way repeated measures analysis of variance (ANOVA) formulation, the methods presented are applicable to genotype data on single nucleotide polymorphisms (SNPs) and *categorical* environmental exposures. Although we study the methods in the context of $G \times E$ or $G \times G$ interactions, they can be used for exploring interactions in any two-way classification array. We then proceed to extend our treatment of the problem to account for individual level repeated measures, beyond the initial cell-means-based approach.

The statistical interaction term, as described by the inclusion of a product term in a regression model, reflects that the effect of the row variable and the column variable may not be additive in their contribution to the quantitative trait. Researchers have described a variety of models on the structure of this non-additivity. Tukey [5] proposed his well-known single degree of freedom (df) test for non-additivity where the interaction is modeled as being proportional to the product of the main effects . Mandel [6] proposed two other more general interaction models where the interaction is proportional to the row main effects or column main effects. Along with these classical models, the newer class of models we explore for repeated measures data is the additive main effects and multiplicative interaction model (AMMI) first introduced by Gollob [7] and then developed by several authors [8–14]. The AMMI models also target towards a sparse representation of interaction terms but not through main effects. This class of models has been used to analyze data from a balanced experimental design to study genotype $\times$ environment in agriculture and crop sciences [15, 16]. Recently, Barhdadi and Dubé applied this class of models to observational studies of gene–gene interaction [17]; Alin and Kurt also contained an overview [18]. All of the aforementioned work were primarily developed for cross-sectional data with fixed effects. Researchers extended the AMMI model to the situation when one of the two factors is fixed, and the other is random, again for cross-sectional data [19–21]. Meyer also considered multiple correlated quantitative traits on the same subject in a mixed model framework [22].

We first focus on developing simple screening tools for interactions with repeated measures longitudinal data based on cell means where both row factor and column factor are considered as fixed effects. The fitting of an AMMI model is based on a singular value decomposition (SVD) of the residual matrix, after removing row and column main effects and retaining the 'leading' (in many cases, the first) component of this representation. The interaction is then represented by the largest characteristic root and the corresponding right and left singular vectors of the interaction matrix, and the remainder terms are attributed to residual noise. Thus, by considering a reduced rank approximation (rank one approximation if only the first component is retained) to the interaction matrix, one is able to save df and enhance efficiency when compared with the saturated interaction model. To this end, we call this method 'principal interactions analysis' (PIA) because of its similarity with 'principal components analysis'. We provide visual/diagnostic tools to isolate subject-specific and time-specific contributions to the principal interaction factors. We carry out a comparative simulation study of the AMMI model with the more traditional models proposed by Tukey and Mandel for repeated measures data, which is not present in the literature. The simulation results illustrate that the AMMI models perform well across a spectrum of scenarios and offer far superior protection against model misspecification than the Tukey/Mandel models. Specifically, for detecting epistasis in the absence of main effects of any of the genetic loci, the AMMI model outperforms models that attempt to parameterize interaction as a function of main effects.

We base the primary investigation in this paper on averaging the response per subject and per cell and then applying the Tukey/Mandel/AMMI models. This approach is appealing in terms of its simplicity, fast computation due to closed-form analytic expression of the test statistics, and can be applied on the basis of simple summaries of the data. However, this approach is certainly limited in terms of its applicability and statistical adequacy if there is varying time effect on the response. The cell-mean approach also fails to properly account for unbalanced nature of the data because of the assumed

homoscedasticity on the error distribution of cell means. To ameliorate this criticism, we propose a more sophisticated analysis that uses individual level data in a mixed-effects regression model setting, followed by a resampling-based test for interaction. Maximum likelihood (ML) and restricted maximum likelihood (REML) estimations under these complex nonlinear correlated outcome models are extremely hard, with non-standard asymptotic distribution theory under the null. We adopt two-step procedures and resampling-based tests to circumvent this problem. As expected, the regression approach using individual level data is more powerful than the cell-mean-based approach and provides right control of type I error under unbalanced designs. For cross-sectional studies, Barhardi and Dubé [18] note that a standard generalized linear model regression with saturated interaction terms that uses individual level data is more powerful than the reduced df tests based on cell-means model. However, they do not propose any alternative to account for unbalanced data or use individual level observations under Tukey/Mandel/AMMI models, even in cross-sectional studies. Thus, the paper makes several original contributions.

The example we consider comes from the Normative Aging Study (NAS), a multidisciplinary longitudinal study of aging in Eastern Massachusetts established by the Veterans Administration in 1963. We consider hearing threshold as measured by pure-tone audiometric examinations at every visit until 1996 as our outcome of interest. Up to eight repeated measures of hearing threshold per subject are available, with 60% subjects having four or more measurements. We explore interplay of two genes, catalase (CAT) and heme-oxygenase 1 (HMOX-1), both involved in the oxidative stress pathway and an occupational noise variable that was derived from lifetime history on job titles. This environmental exposure has five ordinal categories. We illustrate both $G \times G$ and $G \times E$ analysis and investigate changing contribution of the interaction term over time. We compare results of cell-mean-based analysis with resampling tests using individual level data.

We organize the rest of the paper as follows. In Section 2, we describe the three classical models, Tukey's 1-df and Mandel's row and column models. We present the test statistics and the corresponding df/asymptotic null distribution. In Section 3, we discuss the PIA via using the AMMI model. We propose diagnostics for follow-up analysis in terms of subject-specific and time-window-specific contributions to the interaction term. Sections 2 and 3 are based on responses reduced to cell means. In Section 4, we propose analytic approaches that uses individual level data under a mixed-effects regression setup, followed by a resampling-based test statistic for the interaction term. This strategy accounts for unbalanced repeated measures data. In Section 5, we present the data analysis results from the NAS. Section 6 has three parts. Sections 6.1 and 6.2 consider simulation studies on cell-mean-based approaches. Section 6.1 presents simulations to study robustness across the classical interaction models for a general $I \times J$ table. Section 6.2 specifically considers common epistasis models for studying gene–gene interaction [23]. Section 6.3 presents simulation results corresponding to the resampling tests proposed in Section 4, as compared with the cell-means approach. Section 7 concludes with a discussion.

We highlight the new contributions of the paper: (1) application/introduction of the PIA to study $G \times E$ and $G \times G$ effects for repeated measures on quantitative traits; (2) compare with classical interaction models such as Tukey and Mandel's models as applied to repeated measures data; (3) develop visual and diagnostic tools for a better understanding of interaction structures with longitudinally varying outcomes; and (4) introduce novel resampling-based tests for AMMI models (as well as Tukey/Mandel interaction models) that account for unbalanced data structures, uses individual level data under a mixed effects regression modeling framework. The comprehensive simulation studies and the data analyses indicate that the 'PIA' is a promising tool to understand interaction structures and also to trade-off between bias and efficiency in a data-adaptive way under model misspecification.

2. Classical models for interaction

Because the methods are generic to any two-way table, instead of using G and E for the two factors, we use R to denote the row variable with I levels and C to denote the column variable with J levels. Let Y_{hijk} be the hth observation corresponding to the kth subject in the (i, j)th cell of this $(I \times J)$ array. Here, $k = 1, \dots, N$ and $h = 1, \dots, n_k$, with N denoting the total number of subjects and n_k denoting the number of observations corresponding to the kth subject. We consider the following general model:

$$Y_{hijk} = \mu + S_k + R_i + C_j + \gamma_{ij} + \theta_{ik} + \pi_{jk} + \tau_{ijk} + \varepsilon_{hijk}. \tag{1}$$

Here, μ describes the overall mean, R_i and C_j are the row and column main effects, and γ_{ij} describes the interaction between the row and column factors. The standard constraints, $\sum_i R_i = \sum_j C_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$, are placed on the fixed effects parameters. We assume that the possible random effects associated with components $\{S_k\}$ (subject effect), $\{\theta_{ik}\}$ (row × subject), $\{\pi_{jk}\}$ (column × subject), and $\{\tau_{ijk}\}$ (row × column × subject) are jointly normal with zero means and covariance matrix Σ_S. The random errors $\{\varepsilon_{hijk}\}$ are independently and identically distributed with mean zero and variance σ_e^2. A special case of this particular model is the simpler model with only one subject-specific random intercept, namely, S_k, normally distributed with mean zero and variance σ_b^2. We consider this simpler model in our simulation studies and data analyses.

We create a two-way cell-means array, first averaging all observations corresponding to the kth subject in the (i, j)th cell, namely, $\overline{y}_{ijk}$, and then averaging $\overline{y}_{ijk}$, over all subjects in the (i, j)th cell, to obtain $\{\overline{y}_{.ij.}\}$, $i = 1, \ldots I$ and $j = 1, \ldots J$. These cell means will have differing degree of variability, depending on the random effects structure specified in (1) and the number of observations per subject as well as the number of subjects per cell. In a typical observational study, we will certainly have an unbalanced data structure. In the following text, we abuse our notations slightly by dropping the $\{\cdot\}$ suffixes corresponding to subjects and observations corresponding to a subject and describe the models in terms of the $I \times J$ array of cell means $\overline{y}_{ij} = \overline{y}_{.ij.}$.

2.1. A general saturated model for interaction

The implied mean model by (1) for the two-way table in terms of cell means $\overline{y}_{ij}$ is

$$\overline{y}_{ij} = \mu + R_i + C_j + \gamma_{ij} + \overline{\varepsilon}_{ij}, i = 1, \ldots, I, j = 1, \ldots, J, \tag{2}$$

where the interpretation of the fixed effects parameters are as that mentioned previously, but $\overline{\varepsilon}_{ij}$ is the mean of the errors of ε_{hijk} in (1), by first taking averages over errors associated with observations for the kth subject and then over all k subjects within the (i, j)th cell. In the following text, we denote $\overline{y}_{ij}$ by y_{ij}, pretending they represent a single observation corresponding to the $I \times J$ cell [17]. We assume that $\overline{\varepsilon}_{ij} \sim N(0, \tau^2)$. This assumption does not recognize the nonconstant variance in the cell means because of an unbalanced nature of the data. The ML estimates of main effects are then given by the following:

$$\hat{\mu} = y_{..}, \ \hat{R}_i = y_{i.} - y_{..}, \ \hat{C}_j = y_{.j} - y_{...} \tag{3}$$

Let us define the estimated residual contrast after fitting the additive terms as

$$z_{ij} = y_{ij} - \hat{\mu} - \hat{R}_i - \hat{C}_j = y_{ij} - y_{i.} - y_{.j} + y_{...}$$

The df attributed to testing interaction in a saturated model is $(I-1)(J-1)$ and, in that case, $\hat{\gamma}_{ij} = z_{ij}$. With more than one replication per cell, one can test for interaction in a saturated model; however, with a single observation or no replication per cell, one exhausts the df for a saturated interaction model, with no df left for errors. Thus, a test of non-additivity cannot be carried out. In such a situation, researchers have proposed several reduced df tests by imposing special structures on the interaction parameters. These structures can be used for testing interactions in general regression models for a more powerful test with reduced df [24, 25].

2.2. One degree of freedom test for non-additivity [5]

The essential idea behind this model is to think of interaction as $\gamma_{ij} = \theta R_i C_j + \xi_{ij}$, namely, a leading term and some residual noise ξ_{ij} that can be absorbed with the error term ε_{ij}. Thus, of the $(I-1)(J-1)$ df attributed to the interaction term, only one is used to test $H_0 : \theta = 0$ and the rest is attributed to the residual error term, making it possible to test for non-additivity with no replication. Tukey's model is given by the following:

$$y_{ij} = \mu + R_i + C_j + \theta R_i C_j + \varepsilon_{ij}, \tag{4}$$

where θ is the coefficient for the linear-by-linear interaction effect. The least square estimate of θ, denoted as $\hat{\theta}$, is given by

$$\hat{\theta} = \frac{\sum_i \sum_j z_{ij} \hat{R}_i \hat{C}_j}{\sum_i \sum_j \hat{R}_i^2 \hat{C}_j^2} = \frac{\sum_i \sum_j y_{ij} \hat{R}_i \hat{C}_j}{\sum_i \sum_j \hat{R}_i^2 \hat{C}_j^2},$$

where $z_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..}$ is again the estimated residual contrast after removing the additive main effects. This essentially reduces to regressing the cell residuals after fitting the additive terms on the product of the estimated row and column main effects [26]. The model is not identifiable if there are no main effects present as any value of θ yields the same likelihood. Tukey's single df test for non-additivity is obtained by using the test statistic $F = MS_{ll}/MSE$ as presented in Table 1 in the Supplementary material[‡] that has an F distribution with 1 and $(I-1)(J-1)-1$ df under the null hypothesis $H_0 : \theta = 0$.

2.3. Column (row) regression model [6]

Mandel proposed the column regression model and row regression model for testing interactions. In the column regression model, the interaction effect is a linear function of the column main effects, that is,

$$y_{ij} = \mu + R_i + C_j + \lambda_i C_j + \varepsilon_{ij}, \tag{5}$$

where λ_i is the coefficient corresponding to the ith row and $\sum_i \lambda_i = 0$. The ML estimate of λ_i, denoted as $\hat{\lambda}_i$, is

$$\hat{\lambda}_i = \frac{\sum_j z_{ij} \hat{C}_j}{\sum_j \hat{C}_j^2}.$$

The ML estimates of μ and R_i remain unchanged. A test of non-additivity is obtained by constructing an F-statistic for the hypothesis

$$H_0 : \lambda_i = 0, i = 1, \ldots, I.$$

Under the null hypothesis and normality, this test statistic, as described in Table 1 of the Supplementary material, has an F distribution with $(I-1)$ and $(I-1)(J-1)-(I-1)$ df. Table 2 of the Supplementary material presents the ANOVA table for this model. By replacing the columns with the rows, one can equivalently posit a row regression model of the following form:

$$y_{ij} = \mu + R_i + C_j + R_i \eta_j + \varepsilon_{ij}, \tag{6}$$

with $\sum_j \eta_j = 0$ and testing $H_0 : \eta_j = 0, j = 1, \ldots, J$, with the resultant F-statistic having df $\{J-1, (I-1)(J-1)-(J-1)\}$.

Note that models (4)–(6) hierarchically build increasing order of complexity in the interaction structure in a nested manner. For a large two-way array, say a 9×5 array [with saturated interaction df $(I-1)(J-1) = 32$], the interaction tests will have 1 (Tukey's 1-df), 8 (Mandel's column), and 4 (Mandel's row) for the numerator of the F-statistic and 31, 24, and 28 df for the denominator, respectively, thus providing different degrees of efficiency gain and model robustness.

However, all of the aforementioned three models have a particular structure of interaction, specified in terms of main effects. Thus, when there are no main effects, the models encounter problem with likelihood identifiability. Even in the presence of main effects, under any form of misspecification of this specific structure, all of the aforementioned three tests lose tremendous power as discussed in Section 6.

Remark 1 (Tukey's row–column regression model [26])
Tukey extended Mandel's column or row regression model in his seminal paper in 1962 where he introduced the vacuum cleaner strategy for analyzing two-way arrays where a row regression is followed up with a column regression (or vice versa).

$$y_{ij} = \mu + R_i + C_j + \theta R_i C_j + \lambda_i C_j + R_i \eta_j + \varepsilon_{ij}, \tag{7}$$

where λ_i and η_j are the row-specific and column-specific coefficients, with additional constraints $\sum_i \lambda_i = \sum_j \eta_j = 0$ and $\sum_i \lambda_i R_i = \sum_j \eta_j C_j = 0$. The MLEs of μ, R_i, and C_j remain unchanged. The ML estimates of θ, λ_i, and η_j are obtained as follows:

$$\hat{\theta} = \frac{\sum_i \sum_j z_{ij} \hat{R}_i \hat{C}_j}{\sum_i \sum_j \hat{R}_i^2 \hat{C}_j^2}, \qquad \hat{\lambda}_i = \frac{\sum_j z_{ij} \hat{C}_j}{\sum_j \hat{C}_j^2} - \hat{\theta}\hat{R}_i, \qquad \hat{\eta}_j = \frac{\sum_i z_{ij} \hat{R}_i}{\sum_i \hat{R}_i^2} - \hat{\theta}\hat{C}_j. \tag{8}$$

[‡]*Supplementary material may be found in the online version of this article.*

Table 3 of the Supplementary material presents the ANOVA table corresponding to this model. The F-statistic for testing $H_0 : \theta = \lambda_i = \eta_j = 0$, $\forall\ i, j$ under the aforementioned constraints have numerator df $1+(I-2)+(J-2) = (I+J-3)$. The denominator df is thus $(I-1)(J-1)-(I+J-3)$.

For completeness, we provide the description of this more general model but refrain from discussing it any further. For a 3×3 table for $G \times G$ interaction, the Tukey row–column model has df 3, offering little power gain over the saturated model, which has 4 df, another reason for not including this model in our simulation studies.

3. Principal interactions analysis via the AMMI model

Gollob [7] proposed a factor ANOVA model to decompose a two-way table. The essential idea is to represent the $I \times J$ rectangular interaction matrix Γ with interaction parameters γ_{ij} as entries, by the following representation:

$$\Gamma = ADB'.$$

Here, $A = ((\alpha_{im}))$ and $B = ((\beta_{km}))$ are $I \times R$ and $J \times R$ orthonormal matrices ($A'A = B'B = I$) and D is an $R \times R$ diagonal matrix with elements $d_1 \geqslant d_2 \ldots \geqslant d_R$. The maximum rank of Γ is $\min(I-1, J-1)$ because of the sum-zero constraints on the parameters γ_{ij}. This makes the matrix Γ doubly centered. Let $I \leqslant J$, thus the maximal rank of Γ is $I-1$. Let $P = AD^{1/2}$ and $Q = D^{1/2}B'$, then $\Gamma = PQ' = \sum_m p_{im}q_{jm}$, where the p_{im} and q_{jm} satisfy the orthonormalization constraints $\sum_i p_{im}p_{il} = \sum_j q_{jm}q_{jl} = 0$ for $m \neq l$ and $\sum_i p_{im}^2 = \sum_j q_{jm}^2 = 1$ for $m = 1, \ldots, R$.

By this factor representation, for a saturated model, γ_{ij} is perfectly reproduced by,

$$\gamma_{ij} = \sum_{m=1}^{I-1} d_m\alpha_{im}\beta_{jm} = \sum_{m=1}^{I-1} p_{im}q_{jm}.$$

However, one can think of a sparse representation of the interaction matrix by retaining the first $M < I-1$ components of this representation, namely,

$$\gamma_{ij} = \underbrace{\sum_{m=1}^{M} d_m\alpha_{im}\beta_{jm}}_{\text{leading term}} + \underbrace{\phi_{ij}}_{\text{random noise}}$$

This representation gives rise to the following general class of AMMI models [9–12, 14].

$$y_{ij} = \mu + R_i + C_j + \sum_{m=1}^{M} d_m\alpha_{im}\beta_{jm} + \varepsilon_{ij} \qquad (9)$$

$$= \mu + R_i + C_j + \sum_{m=1}^{M} p_{im}q_{jm} + \varepsilon_{ij}. \qquad (10)$$

Eckart and Young [27] showed that, for a fixed M, the least square estimates of (A, B, D), equivalently, $\{\alpha_{im}\}$, $\{\beta_{jm}\}$, and $\{d_m\}$ can be found by expressing the *estimated* matrix $\hat{\Gamma}$ of interaction parameters with entries $\hat{\gamma}_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..}$ in terms of a SVD as specified by the factor model,

$$\hat{\Gamma} = \hat{A}\hat{D}\hat{B}'.$$

An alternative interpretation is that the interaction parameter is expressed as a sum of several successive multiplicative contrasts $\Psi_{Fm} = \sum_i \sum_j \alpha_{im}\beta_{jm}\gamma_{ij}$ such that each contrast is orthogonal to all previous contrasts and accounts for a maximum of the remaining variance. Let $\hat{\Psi}_{Fm}$ denote the estimated normalized orthogonal multiplicative contrast among the interaction parameters $\{\gamma_{ij}\}$ and SS_{Fm} denote the sum of squares due to the mth interaction factor. Then, from classical contrast theory, we know that $SS_{Fm} = \hat{\Psi}_{Fm}^2$. To this end, $\hat{\Psi}_{Fm}$ can be obtained by

$$\hat{\Psi}_{Fm} = \sum_i \sum_j \hat{\alpha}_{im}\hat{\beta}_{jm}\hat{\gamma}_{ij} = \sum_i \sum_j \hat{\alpha}_{im}\hat{\beta}_{jm}y_{ij}.$$

Because $\hat{\Gamma} = \hat{A}\hat{D}\hat{B}'$, we have $\hat{D} = \hat{A}'\hat{\Gamma}\hat{B}$, implying $\hat{d}_m = \sum_i \sum_j \hat{\alpha}_{im}\hat{\beta}_{jm}\hat{\gamma}_{ij}$. So, $\hat{\Psi}_{Fm}$ and $\hat{d}_m$ are equivalent. They both are $\sum_i \sum_j \hat{\alpha}_{im}\hat{\beta}_{jm}\hat{\gamma}_{ij}$. Hence, $SS_{Fm} = \hat{\Psi}_{Fm}^2 = \hat{d}_m^2$. Let SS_{RC} denote the total sum of squares due to row–column interaction. The sum of squares corresponding to the residual interaction after M-successive interaction factors being extracted from $\{\gamma_{ij}\}$ is therefore

$$SS_{Fres} = SS_{RC} - \sum_{m=1}^{M} SS_{Fm} = SS_{RC} - \sum_{m=1}^{M} \hat{d}_m^2 = \sum_{m=M+1}^{I-1} \hat{d}_m^2,$$

as
$$SS_{RC} = \sum_i \sum_j (y_{ij} - y_{i.} - y_{.j} + y_{..})^2 = \sum_i \sum_j \hat{\gamma}_{ij}^2 = \text{Trace}(\hat{\Gamma}'\hat{\Gamma}) = \text{Trace}(\hat{D}'\hat{D}) = \sum_{m=1}^{I-1} \hat{d}_m^2.$$

Table 4 of the Supplementary material contains the ANOVA table corresponding to the AMMI model. The use of pseudo F-tests with various prescriptions for the df is based on heuristic approximations [7, 9]. Essentially, corresponding to the mth interaction factor, there are $I + J + 1$ parameters α_{im}, β_{jm}, and d_m, but there are $2m + 2$ orthonormality constraints due to orthogonality to prior $m - 1$ contrasts and being normalized to unity. Thus, the mth interaction factor has numerator df: $df(m) = (I + J + 1 - (2m + 2))$. The set of M factors together have df $= \sum_{m=1}^{M} df(m)$. The remaining interaction df after fitting first M factors is $(I - 1)(J - 1) - \sum_{m=1}^{M} df(m) = (I - 1 - M)(J - 1 - M)$. Because this pseudo F-test does not have desirable operating characteristics, we relegate the details to Table 4 of the Supplementary material.

A special case of (9) is of particular interest when $M = 1$. Namely,

$$y_{ij} = \mu + R_i + C_j + d_1\alpha_i\beta_j + \varepsilon_{ij},$$
$$\sum_i \alpha_i = \sum_j \beta_j = 0; \quad \sum_i \alpha_i^2 = \sum_j \beta_j^2 = 1. \tag{11}$$

Thus, the test of no interaction is equivalent to testing $H_0 : d_1 = 0$. Johnson and Graybill [10] derived the distributional properties for the likelihood ratio test (LRT) of $H_0 : d_1 = 0$. They show that the ML estimate of d_1, $\hat{d}_1$ say, is given by the square root of the largest characteristic root of $\hat{\Gamma}'\hat{\Gamma}$, say l_1. The maximum value of the likelihood is attained when $\{\alpha_i\}$ and $\{\beta_j\}$ are given by the normalized characteristic vector corresponding to l_1 in $\hat{\Gamma}'\hat{\Gamma}$ and $\hat{\Gamma}\hat{\Gamma}'$, respectively. Consequently, the LRT for $H_0 : d_1 = 0$ versus $H_a : d_1 \neq 0$ is given by

$$\Lambda = \left(\frac{\sum_i \sum_j \hat{\gamma}_{ij}^2 - l_1}{\sum_i \sum_j \hat{\gamma}_{ij}^2} \right)^{IJ/2}, \tag{12}$$

where $l_1 = \hat{d}_1^2$ again is the maximum non-zero (characteristic) root of $\hat{\Gamma}'\hat{\Gamma}$. That is, l_1 is the maximum value of $(\sum_i \sum_j \alpha_i\beta_j y_{ij})^2$ with respect to α_i and β_j subject to the restriction that $\sum_i \alpha_i = \sum_j \beta_j = 0$ and $\sum_i \alpha_i^2 = \sum_j \beta_j^2 = 1$. The critical region for $H_0 : d_1 = 0$ can equivalently be expressed as follows:

$$\Lambda^* = \frac{l_1}{\sum_{m=1}^{I-1} l_m} = \frac{d_1^2}{\sum_{m=1}^{I-1} d_m^2} > \text{constant}.$$

The asymptotic distribution for the LRT statistic is not χ^2. Critical points of Λ^* for several choices of I and J are provided previously [11, 28]. The theory is based on deriving asymptotic property of the ratio of largest characteristic to the trace of a Wishart matrix. We present the details of ML estimation in the Supplementary material.

Remark 2
In general, the number of components M should be chosen in such a way that the residual ϕ_{ij} represents noise and can again be absorbed with ε_{ij} leading to a more powerful test with reduced df. Several studies have investigated cross-validation and significance testing approaches for determining M, the

appropriate number of multiplicative interaction terms to be retained [29–31]. When the aforementioned model is saturated, $M = I - 1$. We focus on the model with $M = 1$ in the remainder of the paper and do not address the issue of data-adaptive selection of M. In our data example, including $M = 1$ component was sufficient.

3.1. Biplot and subject-specific and time-specific contributions

In this section, we describe certain graphical diagnostics to provide insight into interaction structures. In particular, we discuss the best rank-two approximation to an interaction matrix as presented by a biplot [32]. We then introduce diagnostics to assess subject-specific contribution and time varying contribution to the leading interaction term.

3.2. Biplot

The biplot is a graphical planar display of the elements, rows and columns of a matrix. Any matrix of rank two can be displayed as a biplot which is defined through a vector for each row and a vector for each column, such that the inner product represents each matrix element. For a matrix with higher rank, one may use the biplot corresponding to the best rank-two approximation to the original matrix. With the factor analytic representation $\hat{\Gamma} = \hat{A}\hat{D}\hat{B}'$, each entry of the estimated interaction matrix can be approximated by the first two terms of the corresponding factor representation by

$$\hat{\gamma}_{ij} = \hat{d}_1 \hat{\alpha}_{i1} \hat{\beta}_{j1} + \hat{d}_2 \hat{\alpha}_{i2} \hat{\beta}_{j2}.$$

For $G \times G$ interaction, for example, the matrix of interest is a $(I = 3) \times (J = 3)$ matrix with maximal rank $I - 1 = 2$, and this representation is exact. There are several choices of defining the vectors; we define the points $P_i = (\hat{d}_1^{1/2}\hat{\alpha}_{i1}, \hat{d}_2^{1/2}\hat{\alpha}_{i2})$ representing row i and the points $Q_j = (\hat{d}_1^{1/2}\hat{\beta}_{j1}, \hat{d}_2^{1/2}\hat{\beta}_{j2})$ describing column j.

Bradu and Gabriel [33] explained the use of biplots for interaction models. The patterns of the points indicate certain models: additivity (the case of two orthogonal lines), Mandel's row regression model when P_i are collinear and Q_j are scattered, or column regression when Q_j are collinear and P_i are scattered. The AMMI model typically will give rise to a configuration where P_i and Q_j are both scattered. For the special case of AMMI with $M = 1$, the points are not collinear but co-planar on the three-dimensional plane. We use this representation for repeated measures data with the cell-means residual, as described before, to visualize the interaction structure.

3.3. Measures that summarize subject-specific and time-specific contributions of interaction

The question that we stated at the onset was to capture varying effects of time and subject to the interaction term. We take a very different approach than a standard mixed model regression setting as described in the introduction. We utilize the PIA framework and construct measures that summarize variation due to individual differences in the size of the contribution to the leading interaction factors. Variation due to individual differences can be investigated by defining a contrast using the estimated factor weights $(\hat{\alpha}_i, \hat{\beta}_j)$ and the individual person level means for the kth subject in the (i, j)th cell, namely, $y_{.ijk}$, by computing the following N subject-specific regression weights for the mth interaction factor [7]:

$$\hat{d}_{km} = \sum_i \sum_j \hat{\alpha}_{im} \hat{\beta}_{jm} y_{.ijk}, \quad k = 1, \ldots, N.$$

The larger the value $\{d_{km}\}$ for the kth subject, the larger the absolute contribution of the mth factor to determine the subject's mean $y_{.ijk}$. For the mth interaction factor, the variation in the contribution of each individual can be calculated by a squared term, $(\hat{d}_{km} - \hat{d}_{.m})^2$, where $\hat{d}_{.m}$ is $\frac{1}{N}\sum_k \hat{d}_{km}$. The aggregate measure of squared deviation $\sum_{k=1}^{N}(\hat{d}_{km} - \hat{d}_{.m})^2$ captures total subject-specific variability in the mth interaction term around the average regression weight.

Variation of different time contributions can be investigated in a similar manner. We may define T time intervals of a given width w to cover the study period. We calculate $y_{tij.}(t = 1, \ldots, T)$, which is the averaged score of *all* observations in the ith row and the jth column over the w year follow-up

period. Then, the variation due to time can be investigated by computing the T quantities for the mth factor

$$\hat{d}_{tm} = \sum_i \sum_j \hat{\alpha}_{im} \hat{\beta}_{jm} y_{tij.}, \quad t = 1, \ldots, T.$$

We can calculate the relative contribution of each time by using a squared term, $(\hat{d}_{tm} - \hat{d}_{.m})^2$, where $\hat{d}_{.m}$ is now $\frac{1}{T} \sum_t \hat{d}_{tm}$. The aggregate measure of squared deviation $\sum_{t=1}^{T} (\hat{d}_{tm} - \hat{d}_{.m})^2$ captures time variability in the mth interaction term around the average regression weight across the study period. We illustrate these diagnostics and graphical representation in the following section through analyzing data from the NAS.

4. Resampling-based interaction tests using mixed models

In Sections 2 and 3, we discussed treatment of the interaction testing problem in terms of reducing the response in a crude way to average per person and per cell and then thinking of the cell mean as a single observation per cell. Although this approach is fast and simple, it has many limitations such as ignoring the time variation in response and ignoring the unbalanced nature of the study design. However, because of nonlinear structure in the parameters, for example, terms of the form $\theta R_i C_j$ for Tukey's model, $\lambda_i C_j$ for Mandel's column regression model, and $d_1 \alpha_i \beta_j$ in AMMI model, ML estimation and establishing asymptotic theory for a general random effects model are hard. There is very little literature for unbalanced data situations with these models and almost non-existent literature in the repeated measures observational study setting. Maity *et al.* [25] presented the most general treatment of Tukey's model but only considered the testing problem. Meyer [22] considered balanced data but correlated multiple response. Solving the testing problem for each of the Tukey/Mandel/AMMI models for a general random effects structure and unbalanced data, accompanied with analytical asymptotic theory, remains beyond the scope of the paper. In this section, we develop a set of novel resampling-based tests for this class of models, which have not been proposed in the literature. The permutation tests use individual repeated measures and utilize a general mixed effects ANOVA/regression framework followed by a permutation-based null distribution of the test statistic.

4.1. Two-step regression procedure for Tukey/Mandel models

In step 1, we fit a standard saturated interaction model with γ_{ij} using all $(I-1)(J-1)$ df by including product terms of row and column indicators in the model. For example, with a random intercept structure for subject k, we first fit a linear mixed effects model

$$y_{hijk} = \mu + S_k + R_i + C_j + \gamma_{ij} + \varepsilon_{hijk}, \tag{13}$$

where $S_k \sim N(0, \sigma_b^2)$ and $\varepsilon_{hijk} \sim N(0, \sigma_e^2)$. We obtain REML estimates of fixed effects $\hat{\mu}$, $\hat{R}_i$, and $\hat{C}_j$ and variance components $\hat{\sigma}_b^2$ and $\hat{\sigma}_e^2$ under this model. We then construct the marginal residuals:

$$r_{hijk} = y_{hijk} - \hat{\mu} - \hat{R}_i - \hat{C}_j.$$

Recall that even for cross-sectional unbalanced data, there are no closed-form expressions of $\hat{R}_i$ and $\hat{C}_j$ as in the two-way balanced ANOVA. Using dummy variable regression model is the best way to express the model estimates in the unbalanced case. In step 2, residuals from step 1, r_{hijk}, are regressed on $\hat{R}_i \hat{C}_j$,

$$r_{hijk} = \theta \hat{R}_i \hat{C}_j + \varepsilon'_{hijk}.$$

At step 2, we have used compound symmetry covariance structure, but one can allow for a user-defined covariance structure in ε'_{hijk}, depending on assessment of model fit criterion. Note that one can alternatively use the subject-specific residuals from step 1, namely, $r^s_{hijk} = y_{hijk} - \hat{\mu} - \hat{S}_k - \hat{R}_i - \hat{C}_j$, and use them as outcomes in a second-stage regression model. Irrespective of the choice of residuals (marginal or subject specific) that changes the estimation/choice of variance covariance matrix for ε', the estimate of θ at step 2 appears to remain unbiased under the two-step procedure if the original generating model had the structure $\gamma_{ij} = \theta R_i C_j$. Note that we are exploiting the idea that, after removing the additive

term, we are expressing the residual variability attributable to both interaction and random error through a second-step correlated outcome model. To test $H_0 : \theta = 0$, we used a test statistic that has an analogous form to what we used in the cell-means approach, namely, $T_{Tukey} = \hat{\theta}^2/\hat{\sigma}_e^2$. To elicit the null distribution under the two-step approach, we adopt the following resampling strategy. Note that this exercise of simulating the null can be tricky as one would like to still preserve the main effects and simply eliminate the interaction pattern. Permuting the Y values or subjects across cells will remove the interaction but will destroy the main effects structure as well. To bypass this problem, we generate pseudo data $\mathbf{Y}^*$ under $H_0 : \theta = 0$,

$$Y_{hijk}^* = \hat{\mu} + S_k^* + \hat{R}_i + \hat{C}_j + \varepsilon_{hijk}^*, \quad S_k^* \sim N(0, \hat{\sigma}_b^2), \quad \varepsilon_{hijk}^* \sim N(0, \hat{\sigma}_e^2), \tag{14}$$

where $\hat{R}_i, \hat{C}_j, \hat{\sigma}_b^2$, and $\hat{\sigma}_e^2$ are REML estimates obtained from step 1 model. We generate 1000 such pseudo datasets reflecting the null. For each such pseudo dataset (containing N individuals, each with number of repeated measures as recorded in the original dataset), we fit a two-step regression approach exactly as our analysis of original data to compute $T_{Tukey}^* = \hat{\theta}^{*2}/\hat{\sigma}_e^{*2}$. We then compare our observed value of T_{Tukey}^{obs} on the basis of the original data with the sample percentiles of these 1000 test statistics generated under the null to obtain the p-value corresponding to the test statistic.

Similarly, for Mandel's column regression model, we regress the residuals from step 1 saturated interaction model to obtain a set of $I-1$ second-step regression coefficients:

$$r_{hijk} = \lambda_i \hat{C}_j + \varepsilon_{hijk}', \text{with} \qquad \sum_{i=1}^{I} \lambda_i = 0.$$

Again, we consider the familiar form of a multiple of the F-type test statistic to test $H_0 : \lambda_1 = \cdots = \lambda_I = 0$. Compute it for pseudo data $\mathbf{Y}^*$ as in (14) $T_{MandelR}^* = \sum_{i=1}^{I} \hat{\lambda}_i^{*2}/\hat{\sigma}_e^{*2}$. We then compare the observed value of the test statistic with the distribution of the test statistics obtained by analyzing the 1000 pseudo datasets generated under the null. The vast literature on choosing appropriate covariance matrices at the steps 1 and 2 linear mixed effects model can be applied to a particular data analysis as long as the pseudo datasets are generated and analyzed under the same choices. One could also postulate alternative forms of the test statistics instead of the ones we borrowed from the cell-means model.

5. Exploring $G \times G$ and $G \times E$ in the normative aging study

The NAS is a multidisciplinary longitudinal study of aging in Eastern Massachusetts established by the Veterans Administration in 1963 [34]. Data were collected every 3–5 years, including extensive physical examination, laboratory, anthropometric, and questionnaire data. The outcome we consider is hearing threshold as measured by pure tone average of thresholds at frequencies of 0.5, 1, 2, and 4 kHz. Smaller threshold represents better hearing ability [35]. The dataset contained a total of 662 individuals. Each individual had at least two measurements, and 62% of them had at least four measurements over time. Table I shows descriptive characteristics of the study population. We considered two SNPs on genes related to oxidative stress pathway and one environmental exposure, namely, occupational noise. The two genetic markers were rs2071746 (T/A) on HMOX-1, a stress response protein, which may offer protection against oxidative stress, and rs1001179 (C/T) on CAT, a gene that decomposes hydrogen peroxide. Mordukhovich *et al.* have studied both of these SNPs in NAS as an effect modifier in a recent study of black carbon on blood pressure [36]. However, the role of these genetic markers related to oxidative stress defense has not been studied for hearing threshold outcomes. Park *et al.* [37] created an ordinal measure for lifetime exposure to noise with five levels (1 reflecting lowest noise exposure and 5 indicating highest) on the basis of prior literature.

The estimated minor allele frequencies for the SNPs considered on CAT and HMOX-1 were 0.19 and 0.46, respectively and both SNPs were in Hardy–Weinberg equilibrium ($p = 0.30$ and 0.67, respectively). There can be a maximal number of $M = I - 1 = 3 - 1 = 2$ principal interaction factors here, and the biplot representation is exact. The upper panel of Figure 1 shows the cell means corresponding to the $G \times G$ cross-classification, the matrix $\hat{\Gamma}$, and the corresponding SVD along with the corresponding biplot. The plot of cell means suggest evidence for interaction. In the biplot, the points representing the column array appear to be nearly collinear, suggesting possible evidence for Mandel's column regression

Table I. Descriptive characteristics of 662 study participants in the Normative Aging Study considered in our data analysis.

Variable	Mean	SD
PTA hearing threshold (dB) (Y)	10.86	6.54
Age (years)	41.66	8.77
Body mass index (kg/m^2)	25.71	2.76

	N	Percent
Race (white)	645	97.43
Education (>12 years)	381	57.55
Type 2 diabetes	13	1.96
Hypertension	28	4.23
Pack years of cigarettes		
0	205	30.97
<30	336	50.76
≥30	121	18.28
Genes (G)		
CAT (C/T) rs1001179		
CC	403	65.96
CT	179	29.3
TT	29	4.75
HMOX-1 (T/A) rs2071746		
TT	171	27.67
TA	320	51.78
AA	127	20.55
Environment (E)		
Level of noise exposure		
1	120	18.13
2	95	14.35
3	182	27.49
4	153	23.11
5	112	16.92
Number of repeated measures on PTA per subject		
2	129	19.49
3	122	18.43
4	155	23.41
5	147	22.21
6	85	12.84
7	20	3.02
8	4	0.60

Age, body mass index, health status, and smoking variables are measured at baseline. Pure tone average (PTA) hearing threshold is averaged over all repeated measures.

model. Table II presents the results from the different fitted models along with a random intercept mixed model (under a compound symmetry covariance) with main effects of both SNPs and saturated $G \times G$ interaction. The interaction is marginally significant in only Mandel's column regression model where the interaction is assumed to be proportional to the main effect of rs2071746 on HMOX-1 ($p = 0.06$) and not significant in any other model. There is evidence of main effect of HMOX-1 as well in the column regression model ($p = 0.05$) and from the descriptive statistics.

The AMMI model using the LRT with $M = 1$ has a p-value between 0.1 and 0.2 for the leading principal factor, whereas the pseudo F-test [9] used in the AMMI macro in SAS (SAS Institute Inc., Cary, NC, USA) [38] has a much larger p-value of 0.61. The 5% upper critical value of AMMI-LRT [10] for a 3×3 array is 0.9994, whereas our observed value is 0.9533. The leading characteristic root of $\hat{\Gamma}'\hat{\Gamma}$,
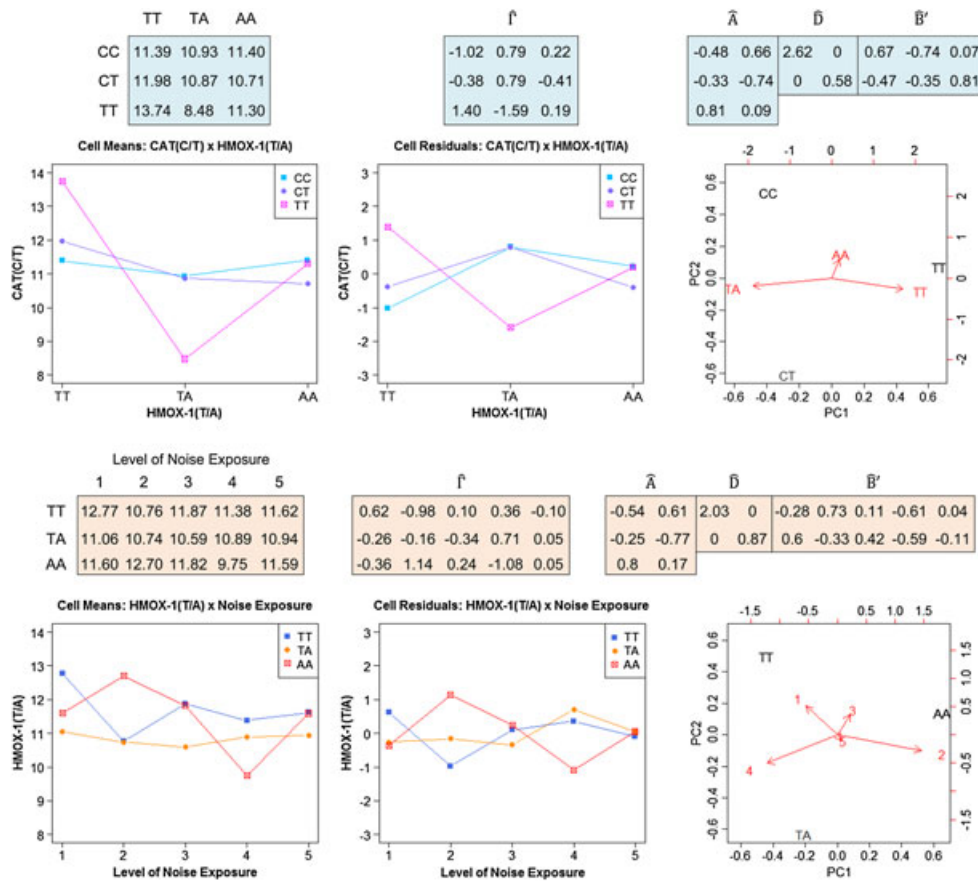
Figure 1. Cell means, residuals after eliminating additive row and column main effects, and the SVD of the estimated $\hat{\Gamma}$ matrix for $G \times G$ (top panel) and $G \times E$ (bottom panel) analyses. The numerical arrays are accompanied by graphical displays of the cell means, entries of $\hat{\Gamma}$, and the biplot representation. Results are based on the Normative Aging Study data.

namely, $l_1 = \hat{d}_1^2$ is 6.82 and $l_2 = \hat{d}_2^2 = 0.33$. Because the LRT statistic also represents the fraction of the total variability due to the interaction term explained by the first component, (LRT $= \hat{d}_1^2/(\hat{d}_1^2 + \hat{d}_2^2)$), we note that the first principal interaction component explains 95% of the interaction sum of squares and the second principal interaction component can be attributed to random noise.

We carried the same analysis for $G \times E$ model with a 3×5 table for HMOX-1 and occupational noise exposure. The maximal number of interaction factors is still $3 - 1 = 2$. The lower panel of Figure 1 displays the cell means corresponding to the $G \times E$ cross-classification, the matrix $\hat{\Gamma}$, and the corresponding biplot. We observed no obvious pattern in the cell-means plot and biplot. Table II also shows the results of fitting different models for the interaction between HMOX-1 and noise exposure and fitting a mixed model with random intercepts. We detected no main effects of gene, exposure, or $G \times E$ interaction in any of the models. The AMMI model using the LRT with $M = 1$ has a *p*-value greater than 0.4 for the leading principal factor, whereas the pseudo F-test used in the AMMI macro in SAS has a larger *p*-value of 0.50. The 5% upper critical value of AMMI-LRT from Johnson and Graybill for a 3×5 array is 0.9648, whereas our observed value is 0.8476. The leading characteristic root of $\hat{\Gamma}'\hat{\Gamma}$, equivalently, $l_1 = \hat{d}_1^2$ is 4.14 and $l_2 = \hat{d}_2^2 = 0.74$. Thus, only the first principal interaction component explains 85% of the interaction sum of squares and the second principal interaction component explains the remaining noise. An LRT based on fitting the two nested models also supports the same conclusion.
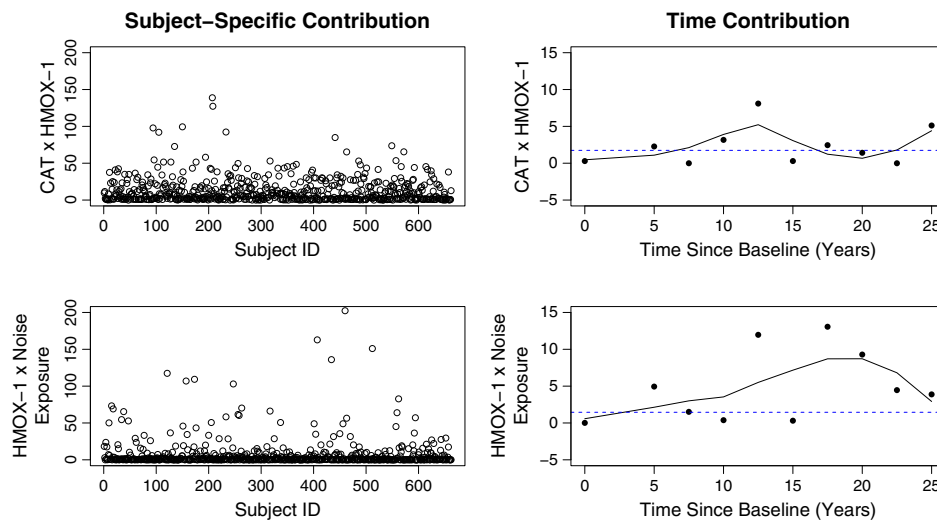
5.1. Subject-specific and time-specific contributions to the principal interaction factors

The left column in Figure 2 displays the contribution of the 662 individuals to the first interaction factor for $G \times G$ and $G \times E$ analysis as computed by the $(\hat{d}_{k1} - \hat{d}_{.1})^2$ term as described in the previous

Table II. Analysis results for gene–gene and gene–environment interactions in the Normative Aging Study.

Model	Hypothesis	Numerator df	F	p-Value (cell mean)	p-Value[†] (resampling)
Analysis results for CAT (C/T) × HMOX-1 (T/A)					
Tukey's 1-df for nonadditivity	$H_0 : \theta = 0$	1	0.87	0.20	0.13
Mandel's row (CAT) regression	$H_0 : \eta_j = 0$	2	0.91	0.52	0.32
Mandel's column (HMOX-1) regression	$H_0 : \lambda_i = 0$	2	14.84	0.06	0.03
AMMI	First PI	3.57	$F^* = 0.11$	0.61	0.41
AMMI	First PI		LRT = 0.9533	$0.1 < P < 0.2$	
Mixed model (random intercept, saturated)	CAT × HMOX-1	4	1.07	0.37	
Analysis results for HMOX-1 (T/A) × noise exposure					
Tukey's 1-df for nonadditivity	$H_0 : \theta = 0$	1	1.06	0.34	0.25
Mandel's row (HMOX-1) regression	$H_0 : \eta_j = 0$	4	0.19	0.93	0.85
Mandel's column (noise) regression	$H_0 : \lambda_i = 0$	2	0.97	0.43	0.25
AMMI	First PI	6.36	$F^* = 1.43$	0.50	0.69
AMMI	First PI		LRT = 0.8476	$p > 0.40$	
Mixed model (random intercept, saturated)	HMOX-1 × noise	8	0.61	0.77	

Two SNPs, rs2071746 on HMOX-1 and rs1001179 on CAT gene, are considered for $G \times G$ analysis. The $G \times E$ analysis considers the interaction between the same SNP on HMOX-1 and occupational noise exposure. We preset results from the four models Tukey's 1-df, Mandel's row, Mandel's column, and PIA via AMMI with one component. The last column presents the results of resampling-based tests as discussed in Section 4. F^* is the pseudo F-value with fractional df [9]. LRT is the likelihood ratio test statistic [10].
[†]Results from resampling-based tests using individual level data.

Figure 2. Subject-specific and time-specific contributions to the first interaction factor in HMOX-1 × CAT (upper panel) and HMOX-1 × occupational noise interaction (lower panel). Each point in the plot presents the squared deviations as described in Section 3. Results are based on the Normative Aging Study data.

section, for $k = 1, \ldots, 662$. One can note that there appears to be more subject-specific variability in the $G \times G$ analysis than the $G \times E$ analysis from this plot. The sum of squared deviations, namely, $\sum_{k=1}^{662} (\hat{d}_{k1} - \hat{d}_{\cdot 1})^2$, has value 7663 for $G \times G$ analysis and 4524 for $G \times E$ analysis. Figure 1 of the Supplementary material presents similar plots corresponding to the second interaction factor that reflects much lesser magnitude of subject-specific variability.

We can investigate variation of different time contributions in a similar manner. We considered 10 time intervals of 2.5 years each to cover the entire study period of 1963–1996. The last time interval contained all observations after 25 years of follow-up. We then calculated $y_{tij\cdot} (t = 1, \ldots, 10)$, which is the averaged score of all observations in the tth 2.5-year follow-up period for all subjects in that (i, j)th cell. Because of the width of the time interval, there was one observation per individual in each interval. The right column in Figure 2 shows the variation due to different time periods in the contribution of the first interaction factor. For the gene–gene interaction (CAT and HMOX-1), time has less varying contribution to the interaction factor as the curve indicates. On the other hand, time had a substantial effect on the interaction between gene HMOX-1 and occupational noise exposure. It appears that the time window around 10–20 years of follow-up shows strongest contribution than the period of 0–10 or 20–25 years. With progressing age, the onset of hearing loss becomes more common and the variation in the quantitative trait is highest in the intervening period. The results suggest that the effect modification of cumulative noise exposure is most relevant in that 'window of vulnerability' where average age of the study subjects were in the age group of 55–65 years. The sum of squared deviations, namely, $\sum_{k=1}^{10} (\hat{d}_{t1} - \hat{d}_{\cdot 1})^2$, has value 23.2 for $G \times G$ analysis and 49.9 for $G \times E$ analysis. Figure 1 of the Supplementary material presents similar plots corresponding to the second interaction factor, showing almost no time-specific variability.

Although these graphical diagnostics do not establish a 'statistical significance' of a $G \times G \times$ Time or $G \times E \times$ time term, they do provide important insight into longitudinal features of the interaction factor. In fact, by fitting a mixed effects model with a compound symmetry error structure, a fixed main effects of G, E, and continuous time, all pairwise interactions between G, E, and time, and $G \times E \times$ time, the three-way term is highly significant with $p < 10^{-3}$.

We also used the resampling-based tests described in Section 4 that uses individual observations to explore $G \times G$ and $G \times E$ effects in the NAS data. The HMOX-1 $\times$ CAT interaction is significant in Mandel's column regression model ($p = 0.03$). For $G \times G$ analysis, the AMMI ($M = 1$) model using permutation test with Gollob's statistic has a p-value of 0.41 for the leading principal interaction factor. The observed value of the two characteristic roots of $\hat{\Gamma}'\hat{\Gamma}$, namely, $l_1 = \hat{d}_1^2 = 6.27$, $l_2 = \hat{d}_2^2 = 0.38$ based on the SVD of $\hat{\Gamma}$ under a saturated interaction model. Thus, the first interaction factor contributes 94% of the total contribution of the interaction term. On the other hand, no significant interaction is detected for $G \times E$ interaction analysis for HMOX-1 and occupational noise exposure. The AMMI model using resampling test has a p-value of 0.69 for the leading principal interaction factor. The observed value of $l_1 = \hat{d}_1^2 = 3.83$, $l_2 = \hat{d}_2^2 = 0.48$. Thus, the first interaction term explains 89% of variability attributed to interaction term.

6. Simulation study

We carried out a simulation study to assess the power and type I error properties of the four tests for interaction (Tukey's 1-df, Mandel's row and column, and AMMI-LRT with $M = 1$). We also considered common epistasis models beyond these four models. We generated individual level data on outcome Y with $n_k = 4$ repeated measures on each subject k for a total of N subjects. The general description of the model, following the notations of Section 2, is given by,

$$Y_{hijk} = \mu + S_k + R_i + C_j + \gamma_{ij} + \varepsilon_{hijk}, \tag{15}$$

with the error $\varepsilon_{hijk} \sim N(0, \sigma_e^2)$, the subject-specific random intercepts $S_k \sim N(0, \sigma_b^2)$, and $\{\varepsilon, S\}$ are mutually independent. Thus, the correlation between any two observations on the same subject is given by $\rho = \sigma_b^2/(\sigma_e^2 + \sigma_b^2)$. We changed the structure of γ_{ij} according to the different simulation models. We first generated the cell means and then the vector of observations per individual with given mean and covariance structure from a multivariate normal distribution.

Section 6.1 presents simulation design and results when the data are generated from each of the four interaction models for a general $I \times J$ table. We consider a 3×3 and a 9×5 setting. Section 6.2 specifically focuses on simulation under common epistasis models [23, 39, 40] for studying gene–gene interaction

(thus, 3×3 tables) with repeated measures on quantitative traits. In all analyses in Sections 6.1 and 6.2, we summarized data by first computing person level average and then by taking the average over all individuals in each cell. We then fitted the four models under consideration and implemented tests for interaction as described in Sections 2 and 3. Under each simulation setting, we generated 1000 datasets, each with N individuals and each individual having four repeated measures. We recorded the percentage of rejections for the null hypothesis of no interaction. For evaluating the type I error, we generated data under the additive model. We considered two settings regarding the variance components: $\sigma_e^2 = 4, \sigma_b^2 = 1$ and $\sigma_e^2 = 4, \sigma_b^2 = 4$, leading to $\rho = 0.2$ and 0.5, respectively. We considered $N = 900, 1800, 3600, 7200$ but only present results for $N = 3600$ in the main text, as the relative performance of the tests remains the same across all sample sizes, only the absolute power increases or decreases with increase/decrease in sample size. The Supplementary material contains the results under some additional settings.

Section 6.3 presents simulation results that compare the cell-mean-based models of Sections 2 and 3 with the resampling tests from Section 4. Because the Section 4 resampling tests have more power as a result of using individual level data, to obtain variation in the power curves, we use the same parameter/effect size setting as in Section 6.1 but increase the variance component values to $\sigma_e^2 = 8, \sigma_b^2 = 2$ and $\sigma_e^2 = 8, \sigma_b^2 = 8$.

### 6.1. Simulation under the general two-way interaction models

6.1.1. Design and parameter setting. We simulated data according to each of the four interaction models with the parameters satisfying the constraints described in Sections 2 and 3: Tukey's 1-df, Mandel's row, Mandel's column, and AMMI ($M = 1$). Under each of the four models, for a 3×3 table, the interaction terms were scaled in such a way that they contributed to 15% of the total variation explained by the model, whereas the remainder is attributed to row and column main effects. We describe the specific details of the parameter setting for the 3×3 table the Supplementary material.

While simulating data under the AMMI model, we assigned the entire contribution due to interaction effect to the first interaction factor. We simulated cell frequencies as if we had two unlinked causal loci with allele frequency 0.3 and 0.4 for all 3×3 tables. For the larger 9×5 table, we pretended as if we are considering combinations of the two loci with allele frequency 0.3 and 0.4, respectively, along with an environmental exposure with five categories with prevalence 0.2 in each category. For the 9×5 table, interaction terms were scaled to contribute 20% of the total variability explained by the model, whereas the rest was attributed to main effects. For simulation under the AMMI model, 75% of the variation due to interaction was attributed to the first component in the 9×5 case. We describe the specific parameter setting for the 9×5 table in the Supplementary material.

6.1.2. Main results. The header on each Figure 3 states the true simulation model while all four 'test' models are fitted under each simulation scenario. The left panel in Figure 3 shows the simulation results corresponding to four tests for a 3×3 table. When the true model is Tukey's 1-df, surely Tukey's 1-df is the most powerful test (100% for $\sigma_b^2 = 1, 4$). Mandel's row and Mandel's column models, being more general than Tukey's 1-df, can capture the interaction structure as well. AMMI is the worst in this setting but has a power around 33% for $\sigma_b^2 = 1$ and around 21% for $\sigma_b^2 = 4$. For simulation under Mandel's row regression model, Mandel's row model obviously has highest power (98% and 85% for $\sigma_b^2 = 1, 4$ respectively), whereas Tukey's 1-df and Mandel's column model can not detect any interaction and has zero power. Again AMMI is less powerful, with power 25% and 16% for $\sigma_b^2 = 1, 4$ respectively. Similar feature holds for Mandel's column model where Tukey's 1-df and Mandel's row fail completely with zero power but AMMI still can capture some interactions (AMMI: 18% and 11% for $\sigma_b^2 = 1, 4$ respectively). With AMMI as the simulation model, all other alternatives fail to capture the interaction in the 3×3 setting except the true model. Note that power decreases as σ_b^2 increases in all cases.

The right panel in Figure 3 presents simulation results for the 9×5 array. The same pattern as described for the 3×3 remain except for the case of Tukey's row–column as the simulation model. For this larger array, Tukey 1-df, Mandel's row, and Mandel's column can capture interactions that are generated by Tukey's row–column model, so does AMMI to a lesser extent. All models fail when data is generated under a general pattern under the AMMI model. Tables 4 and 5 of the Supplementary material present the numerical percentage of rejected null hypotheses corresponding to Figure 3.

To assess the false positive rates or type I error in the absence of interaction, we generated data with only additive main effects and no interaction with $N = 1800, 3600$. Figure 4 presents the percentage of
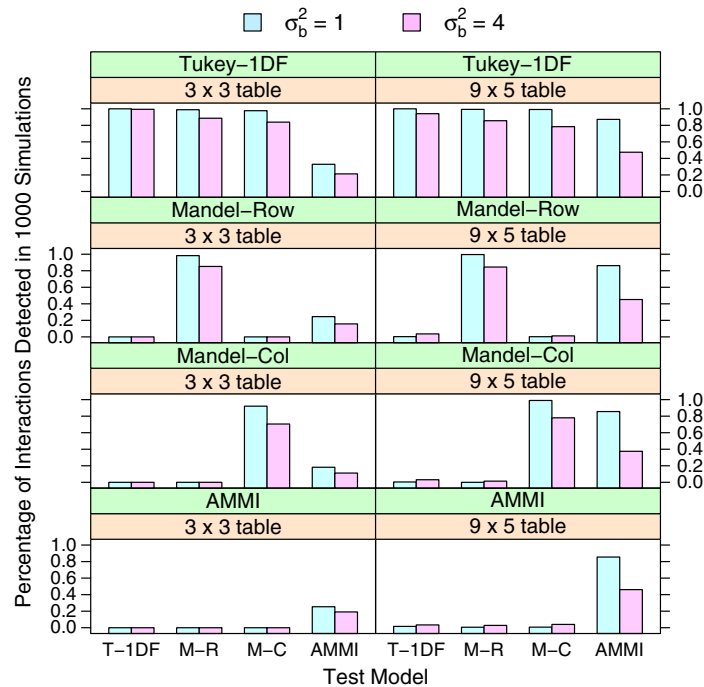
Figure 3. Percentage of interactions detected (or null hypotheses of no interaction rejected) by each of the four tests in the simulation settings corresponding to 3×3 and 9×5 array from 1000 simulated datasets with $N = 3600$. Details are described in Section 6.1. The top label within each box represents the true simulation model, whereas the horizontal axis labels indicate the tests carried out. The error variance σ_e^2 is set at 4 in all cases. Results are based on cell-means model.
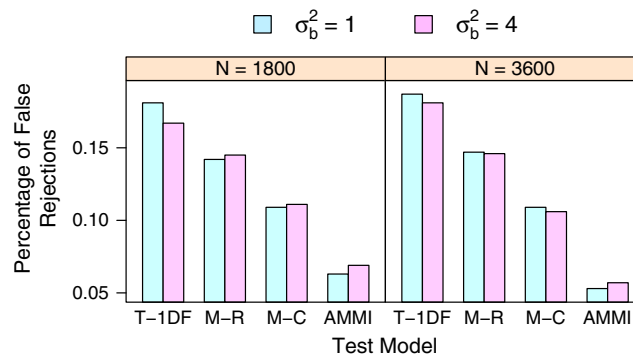
Figure 4. Empirical estimates of type I error rates corresponding to the four interaction tests in a 3×3 array setting based on cell means. Data are generated under additive model, which has only main effects and the set of tests applied to 1000 simulated datasets under each setting. Simulation settings are described in Section 6.1.

false rejections from 1000 simulated datasets at 5% significance level. All type I error rates are inflated than the nominal 5%, especially Tukey's 1-df model. This is due to the use of the cell-mean-based model and ignoring the unbalanced nature of the data. Note that because of asymmetry in genotype frequency, the type I error inflation levels for Mandel's row and column models are not symmetric.

As a summary, the AMMI model follows the 'mediocrity' principle of not being the best but performs reasonably across a spectrum of general interaction models, a robustness feature that is desirable in agnostic search for interaction. None of the other four models possess this robustness property according to our simulation study.

6.2. Simulation under common epistasis models

6.2.1. Design and parameter setting. To evaluate the performance of these five models for studying plausible structures of gene–gene interaction in 3×3 tables with repeated measures on quantitative

traits, we simulated data according to 10 general epistasis models [17]: (1) dominant or dominant (Dom or Dom); (2) dominant or recessive (Dom or Rec); (3) modified model; (4) dominant and dominant (Dom and Dom); (5) recessive or recessive (Rec or Rec); (6) dominant and recessive (Dom and Rec); (7) recessive and recessive (Rec and Rec)[(1)–(7) from [23]]; (8) checkerboard; (9) additive and additive (Add and Add) [(8) and (9) from [40]]; and (10) a general model. The general model has an arbitrary interaction pattern, which was simulated without main effects. The left panel in Figure 5 presents a visual representation of the interaction pattern with true cell means overlaid. In all epistasis models, the grand mean was set to 12. Minor allele frequencies for the two loci are still set at 0.3 and 0.4, respectively.

6.2.2. Main results. Figure 5 shows the results. Tukey's 1-df model and Mandel's row and column models perform well for epistasis models with main effects (1)–(8). Tukey's 1-df model and Mandel's models are substantially more powerful at detecting interactions in model (1)–(8) than the AMMI model. When the main effects do not exist (models 9 and 10 represented in the first row), the AMMI model is the only model that can detect interaction. Thus, in situations where there may not be any main effect of either loci, the AMMI model is able to capture the interaction as it is more flexible than the other four contenders that parameterize interaction in terms of main effects.

Thus to conclude, the AMMI model or performing the PIA does not appear to be a desirable choice for common epistasis structures when compared with Tukey's 1-df, Mandel's row, and Mandel's column models except for the case when there is no main effects of either loci, but epistasis is present.

6.3. Simulation to evaluate the resampling tests

Because the primary goal of the paper is to introduce screening tools in terms of cell-means approach, we conducted limited simulations to compare the performance of the resampling-based tests introduced in Section 4 with the ones in Sections 2 and 3. We considered $N = 1800$ and the 3×3 parameter settings described in Section 6.1. Because tests of interaction accounting for individual observations have much greater power than the tests using cell means, we increased the magnitude of the variance components so that performances of different models can be distinguished. We generated data under two settings: (1) $\sigma_b^2 = 2$ and $\sigma_e^2 = 8$; and (2) $\sigma_b^2 = 8$ and $\sigma_e^2 = 8$, both with $N = 1800$. The within-subject
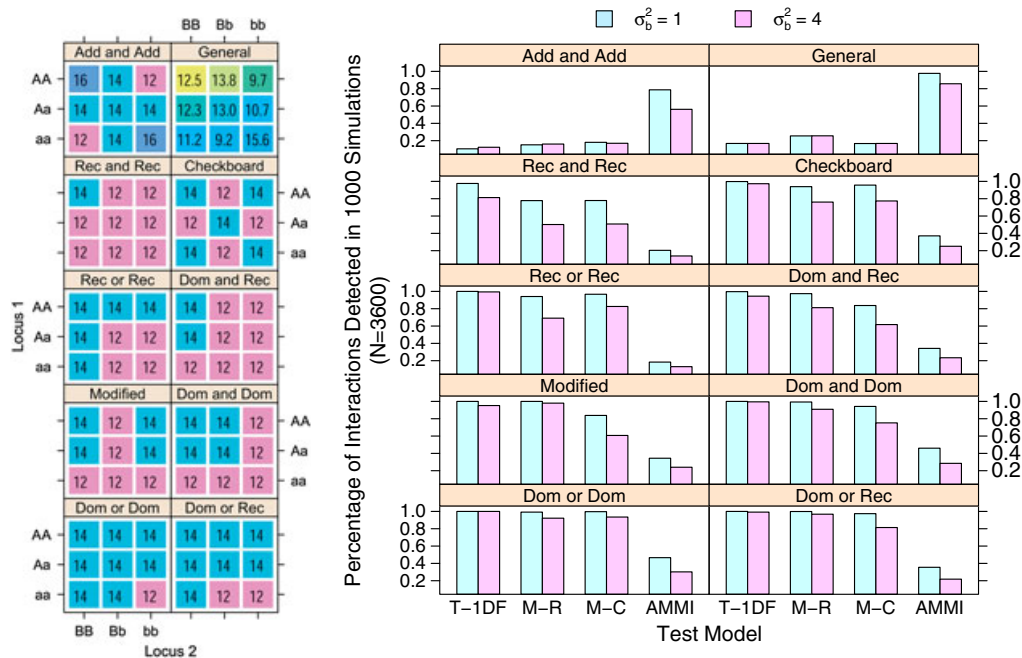
Figure 5. Number of interactions detected (or null hypotheses of no interaction rejected) by each of the four tests in 1000 simulated datasets under 10 common epistasis models. The true models with cell means are displayed in different colors on the left hand panel. The top label within each box represents the true simulation model, whereas the horizontal axis labels indicate the tests carried out. The error variance σ_e^2 is set at 4 in all cases. Simulation settings are described in Section 6.2.

correlations were still 0.2 and 0.5, respectively. Table III shows the power comparison of each of the four interaction models (Tukey's 1-df, Mandel's row, Mandel's column, and AMMI model) to detect interactions based on cell-mean approach and the resampling-based testing approach using individual data.

6.3.1. Main results. As we transition from cell-mean-based model to individual data regression models, all power values generally tend to increase. Especially the power gain for the AMMI model is quite impressive. When the true model is Tukey's 1-df, all models perform reasonably well with $\sigma_b^2 = 2$ (we detected over 80% interactions). With an increase in σ_b^2 to 8, the powers decline and range from 48–60%. For simulation under Mandel's row regression model, the AMMI can detect the interaction, whereas Tukey's 1-df and Mandel's column model have low power (17%, 18% and 10%, 14% for $\sigma_b^2 = 2, 8$ respectively). Similar feature holds for Mandel's column model where Tukey's 1-df and Mandel's row can hardly detect the interaction, but the AMMI maintains 94% power for $\sigma_b^2 = 2$ and 58% power for $\sigma_b^2 = 8$. With the AMMI as the simulation model, Tukey's 1-df model can hardly detect the interaction (28% for $\sigma_b^2 = 2$), whereas Mandel's row and column models have power 49% and 60%, respectively, with $\sigma_b^2 = 2$. The AMMI has power 89% in this case, which is expected as it is the true generation model.

Table III. Comparison of estimated power (percentage of significant interactions detected in 1000 simulations) of resampling-based tests accounting for repeated measures with that of F-statistic-based test using cell means ($N = 1800, \sigma_e^2 = 8$).

True/test model	Cell means				Repeated measures			
	T-1	M-R	M-C	AMMI	T-1	M-R	M-C	AMMI
$\sigma_b^2 = 2$								
Tukey-1df	90.8	70.7	65.0	18.8	89.0	89.0	90.2	86.1
Mandel-Row	0.0	65.0	0.0	12.5	17.4	93.4	18.4	91.8
Mandel-Col	0.0	0.0	50.6	10.0	9.4	9.7	94.7	93.9
AMMI ($M = 1$)	0.3	0.0	0.0	12.9	28.2	49.1	62.6	89.4
Additive *	13.7	10.6	8.0	4.8	3.6	3.9	4.2	4.1
$\sigma_b^2 = 8$								
Tukey-1df	69.8	42.7	39.5	9.9	48.8	54.5	59.0	51.0
Mandel-Row	1.3	39.1	0.2	7.5	10.3	55.4	14.0	51.1
Mandel-Col	0.5	0.1	33.4	8.0	9.5	12.4	65.7	58.1
AMMI ($M = 1$)	2.7	1.3	1.7	10.2	11.3	25.7	29.6	53.1
Additive *	18.0	15.4	12.6	5.6	3.1	5.2	5.8	6.0

*Only main effects but no interaction effect.
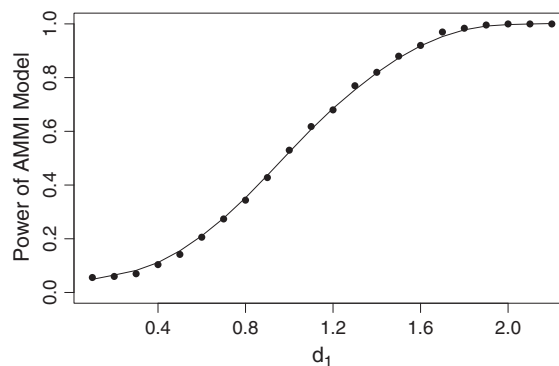
Figure 6. Estimated power of the AMMI ($M = 1$) resampling tests based on individual repeated measures with $d_1 \in (0.1, 2.2)$ under $\sigma_b^2 = \sigma_e^2 = 8$ in a 3×3 array from 1000 simulated datasets with $N = 1800$. The simulation settings are described in Section 6.3.

To assess the false positive rates or type I error of the resampling-based tests, we generated data with only additive main effects without interaction. The last row (additive) presents the number of false rejections from 1000 simulated datasets at 5% significance level. When $\sigma_b^2 = 2$, all type I error rates are maintained at the nominal 5% for the resampling-based tests. Note that the type I error for the cell-mean-based models again are inflated under the additive null.

We investigated in more detail the power curve of the AMMI ($M = 1$) model with repeated measures data to repeated measures with various d_1 values. Figure 6 displays the power curve of AMMI where d_1 ranges from 0.1 to 2.2 under $\sigma_b^2 = \sigma_e^2 = 8$ in a 3×3 array setting with 1000 simulated datasets with $N = 1800$. One can notice that the repeated measures AMMI test is a valid test, maintaining nominal error rate and reasonable power across plausible alternatives.

7. Discussion

In this paper, we have made an initial attempt to explore the idea of principal interaction analysis for repeated measures data on quantitative traits. We compared the proposed approach with other alternative reduced df tests for interaction and established robustness properties of the AMMI model via simulation studies across a spectrum of general interaction models. Our simulation study indicates that the AMMI test may not be very powerful for common epistasis models unless epistasis occurs in the absence of main effects. In our data analyses, we have provided graphical diagnostics to visualize the time-specific and subject-specific contributions to interaction terms.

We have concentrated on the AMMI model with $M = 1$ and used the LRT [10]. We have downplayed the issue of formal selection of the number of interaction components M given the limited scope and length of the paper. That is an interesting question in itself that requires further research and appropriate strategies for inference.

We have primarily adopted a different and somewhat naive route of summarizing the data in terms of cell means in the $I \times J$ configuration and apply classical interaction models for testing non-additivity that are designed for single observations per cell. However, we then have developed new resampling-based tests that used a mixed effects regression framework and fully capitalize on the repeated measures data structure, account for unbalanced data structure in Section 4 for all the models we considered. Our simulation study indicates that the resampling-based tests are valid tests maintaining nominal error levels and have substantially increased power over the cell-mean-based approach, especially when σ_b^2 is large. One can incorporate complex covariance structures, time-varying exposure, longitudinal effects of time, and adjust for covariates by extending the first-step mixed effects regression model in Section 4. One can explore using generalized estimating equation instead of mixed models in Section 4. We have focused on testing; estimation-related properties of these procedures need to be studied as well.

A proper ML approach with repeated measures data and an unbalanced design setting will be more appealing if closed-form expressions for test statistics and their analytic distribution could be obtained instead of the resampling-based approach. Viele and Srinivasan [41] adopted a Bayesian methodology to bypass the complexities to fit the AMMI under complex/unbalanced data structure. One can also fit the AMMI model in a restricted ML framework for mixed models [21, 22]. Tests developed for Tukey's model under a general regression setup (including nonlinear models) can be extended to the more general Mandel's row/column and Tukey 's row–column models [24, 25].

The cell-mean-based approach can be viewed as a screening tool to obtain exploratory/preliminary idea about the interaction structure. In that sense, the PIA for this problem is not just the AMMI test but the accompanying simple visuals and diagnostics as well. The idea of first fitting additive terms and then representing the residual matrix via a sparse decomposition appears to be a promising approach to study non-additivity. Further development of ML-based or REML-based estimation approaches with proper asymptotic theory are warranted to follow-up the current study.

Acknowledgements

References

1. Khoury M, Wacholder S. Invited commentary: from genome-wide association studies to gene-environment-wide interaction studies–challenges and opportunities. *American Journal of Epidemiology* 2009; **169**(2):227.
2. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer Verlag: New York, 2009.
3. Fitzmaurice G, Laird N, Ware J. *Applied Longitudinal Analysis*. Wiley-IEEE: New Jersey, 2004.
4. Lin X, Zhang D. Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 1999; **61**(2):381–400.
5. Tukey J. One degree of freedom for non-additivity. *Biometrics* 1949; **5**(3):232–242.
6. Mandel J. Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association* 1961; **56**(296):878–888.
7. Gollob H. A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika* 1968; **33**(1):73–115.
8. Mandel J. The partitioning of interaction in analysis of variance. *Journal of Research of the National Bureau of Standards-B. Mathematical Sciences* 1969; **73**(4):309–328.
9. Mandel J. A new analysis of variance model for non-additive data. *Technometrics* 1971; **13**(1):1–18.
10. Johnson D, Graybill F. An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association* 1972; **67**(340):862–868.
11. Johnson D, Graybill F. Estimation of σ 2 in a two-way classification model with interaction. *Journal of the American Statistical Association* 1972; **67**(338):388–394.
12. Corsten L, Eijnsbergen A. Multiplicative effects in two-way analysis of variance. *Statistica Neerlandica* 1972; **26**(3): 61–68.
13. Hegemann V, Johnson D. On analyzing two-way AoV data with interaction. *Technometrics* 1976; **18**(3):273–281.
14. Marasinghe M, Johnson D. A test of incomplete additivity in the multiplicative interaction model. *Journal of the American Statistical Association* 1982; **77**(380):869–877.
15. Cornelius P. Statistical tests and retention of terms in the additive main effects and multiplicative interaction model for cultivar trials. *Crop science* 1993; **33**(6):1186–1193.
16. Piepho H. Robustness of statistical tests for multiplicative terms in the additive main effects and multiplicative interaction model for cultivar trials. *TAG Theoretical and Applied Genetics* 1995; **90**(3):438–443.
17. Barhdadi A, Dubé MP. Testing for gene-gene interaction with AMMI models. *Statistical Applications in Genetics and Molecular Biology* 2010; **9**(1).
18. Alin A, Kurt S. Testing non-additivity (interaction) in two-way ANOVA tables with no replication. *Statistical Methods in Medical Research* 2006; **15**(1):63.
19. Oman S. Multiplicative effects in mixed model analysis of variance. *Biometrika* 1991; **78**(4):729.
20. Gogel B, Cullis B, Verbyla A. REML estimation of multiplicative effects in multienvironment variety trials. *Biometrics* 1995; **51**(2):744–749.
21. Piepho H. Analyzing genotype-environment data by mixed models with multiplicative terms. *Biometrics* 1997; **53**(2): 761–766.
22. Meyer K. Factor-analytic models for genotype x environment type problems and structured covariance matrices. *Genetics Selection Evolution* 2009; **41**(1):21.
23. Jung J, Sun B, Kwon D, Koller D, Foroud T. Allelic-based gene-gene interaction associated with quantitative traits. *Genetic Epidemiology* 2009; **33**(4):332–343.
24. Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *The American Journal of Human Genetics* 2006; **79**(6):1002–1016.
25. Maity A, Carroll R, Mammen E, Chatterjee N. Testing in semiparametric models with interaction, with applications to gene–environment interactions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2009; **71**(1):75–96.
26. Tukey J. The future of data analysis. *The Annals of Mathematical Statistics* 1962; **33**(1):1–67.
27. Eckart C, Young G. The approximation of one matrix by another of lower rank. *Psychometrika* 1936; **1**(3):211–218.
28. Hanumara R, Thompson W. Percentage points of the extreme roots of a Wishart matrix. *Biometrika* 1968; **55**(3):505.
29. Gauch Jr H. Model selection and validation for yield trials with interaction. *Biometrics* 1988; **44**(3):705–715.
30. Gauch H, Zobel R. Predictive and postdictive success of statistical analyses of yield trials. *TAG Theoretical and Applied Genetics* 1988; **76**(1):1–10.
31. Piepho H. On tests for interaction in a nonreplicated two-way layout. *Australian & New Zealand Journal of Statistics* 1994; **36**(3):363–369.
32. Gabriel K. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* 1971; **58**(3):453.
33. Bradu D, Gabriel K. The biplot as a diagnostic tool for models of two-way tables. *Technometrics* 1978; **20**(1):47–68.
34. Bell B, Rose C, Damon A. The Veterans Administration longitudinal study of healthy aging. *The Gerontologist* 1966; **6**(4):179.
35. Cruickshanks K, Wiley T, Tweed T, Klein B, Klein R, Mares-Perlman J, Nondahl D. Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. *American Journal of Epidemiology* 1998; **148**(9):879.
36. Mordukhovich I, Wilker E, Suh H, Wright R, Sparrow D, Vokonas P, Schwartz J. Black carbon exposure, oxidative stress genes, and blood pressure in a repeated-measures study. *Environmental Health Perspectives* 2009; **117**(11):1767.
37. Park S, Elmarsafawy S, Mukherjee B, Spiro III A, Vokonas P, Nie H, Weisskopf M, Schwartz J, Hu H. Cumulative lead exposure and age-related hearing loss: the VA normative aging study. *Hearing Research* 2010; **269**(1-2):48–55.
38. Lee E. Statistical analysis software for multiplicative interaction models, 2004. *Unpublished Doctoral Dissertation*.

39. Nothnagel M. Simulation of LD block-structured SNP haplotype data and its use for the analysis of case-control data by supervised learning methods. *The American Journal of Human Genetics* 2002; **71**(suppl 4):A2363.
40. Culverhouse R, Klein T, Shannon W. Detecting epistatic interactions contributing to quantitative traits. *Genetic Epidemiology* 2004; **27**(2):141–152.
41. Viele K, Srinivasan C. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback-Leibler Information. *Journal of Statistical Planning and Inference* 2000; **84**(1-2):201–219.