Published in final edited form as:
Stat Med. 2015 January 15; 34(1): 118–130. doi:10.1002/sim.6323.

GEE for GWAS using Longitudinal Phenotype Data

Colleen M. Sitlani[a,*], **Kenneth M. Rice**[b], **Thomas Lumley**[c], **Barbara McKnight**[b], **L. Adrienne Cupples**[d], **Christy L. Avery**[e], **Raymond Noordam**[f,g], **Bruno H.C. Stricker**[g], **Eric A. Whitsel**[h], and **Bruce M. Psaty**[i,j]

[a]Department of Medicine, University of Washington, Seattle, WA [b]Department of Biostatistics, University of Washington, Seattle, WA [c]Department of Statistics, University of Auckland, Auckland, NZ [d]Department of Biostatistics, Boston University, Boston, MA [e]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC [f]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, NL [g]Department of Epidemiology, Erasmus Medical Center, Rotterdam, NL [h]Departments of Epidemiology and Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC [i]Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, WA [j]Group Health Research Institute, Group Health Cooperative, Seattle, WA

Abstract

Many longitudinal cohort studies have both genome-wide measures of genetic variation and repeated measures of phenotypes and environmental exposures. Genome-wide association study analyses have typically used only cross-sectional data to evaluate quantitative phenotypes and binary traits. Incorporation of repeated measures may increase power to detect associations, but also requires specialized analysis methods. Here we discuss one such method – generalized estimating equations (GEE) – in the contexts of analysis of main effects of rare genetic variants and analysis of gene-environment interactions. We illustrate the potential for increased power using GEE analyses instead of cross-sectional analyses. We also address challenges that arise, such as the need for small-sample corrections when the minor allele frequency of a genetic variant and/or the prevalence of an environmental exposure is low. To illustrate methods for detection of gene-drug interactions on a genome-wide scale, using repeated measures data, we conduct single-study analyses and meta-analyses across studies in three large cohort studies participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium – the Atherosclerosis Risk in Communities (ARIC) study, the Cardiovascular Health Study (CHS), and the Rotterdam Study (RS).

Keywords

GWAS; longitudinal data; gene-environment interaction; rare genetic variants; GEE

[*]Correspondence to: CHRU, University of Washington, 1730 Minor Ave, Suite 1360, Box 358085, Seattle, WA 98101. csitlani@u.washington.edu.

1. Introduction

In recent years, many longitudinal cohort studies have measured genome-wide genetic variation in their participants. Even though repeated measurements of quantitative phenotypes and binary traits are available in these cohorts, genome-wide association studies (GWAS) have largely focused on evaluation of associations at a single point in time. Such cross-sectional analyses generally have lower power than corresponding longitudinal analyses, and they do not permit evaluation of associations with change over time. On the other hand, longitudinal analyses require more specialized and often more resource-intense analysis methods that account for correlation in repeated measurements [1]. Challenges related to missing data, feedback loops, and alignment of time across cohorts further complicate longitudinal analyses.

Despite the challenges associated with longitudinal analyses, they may be particularly useful in low-power settings, such as genome-wide searches for gene-environment interactions [2] or genome-wide analyses of rare variants [3]. The goal of this manuscript is to discuss one option for implementing longitudinal GWAS analysis – generalized estimating equations (GEE) – and to address common challenges that arise.

We will illustrate the use of GEE in the context of an ongoing genome-wide investigation of gene-drug interactions in the pharmacogenetics working group of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium [4]. In this manuscript we focus on a quantitative phenotype – QT interval duration (QT, ms) on the resting, standard twelve-lead electrocardiogram (ECG) – as the outcome, with several different drug classes, including thiazide diuretics, sulfonylurea anti-diabetic agents, tricyclic and tetracyclic antidepressants, and drugs associated with QT-prolongation, as the environmental exposures of interest. Analyses primarily use data from one participating cohort – the Cardiovascular Health Study (CHS) – a longitudinal cohort study with yearly follow-up over a ten-year period from 1989–1999 [5], with additional data from the Atherosclerosis Risk in Communities (ARIC) Study [6] and the Rotterdam Study (RS) [7, 8] used to illustrate methods for meta-analysis.

In section 2, we begin by specifying models of interest and reviewing GEE methods. We then discuss computational challenges in implementing longitudinal GWAS, and we provide illustrations of potential power gains when using longitudinal data instead of cross-sectional data. Finally we suggest solutions for methodological challenges associated with small effective sample sizes and meta-analyses across cohorts. In Section 3, we apply these methods in the context of CHS and the CHARGE pharmacogenetics working group. In Section 4, we summarize our recommendations for longitudinal GWAS and discuss some unresolved challenges that remain.

2. Methods and Simulations

2.1. Statistical Models

The most commonly used mean model in cross-sectional GWAS of unrelated individuals defines exposure additively, based on single-nucleotide polymorphism (SNP) dosage, which may be either observed or imputed [9, 10], and uses link functions as in generalized linear models (GLMs) [11]:

$$g(E[Y_i])=\alpha_0+\alpha_G G_i+\boldsymbol{\gamma}\boldsymbol{Z}_{\mathbf{i}} \quad (1)$$

where i indexes participants, Y is an outcome of interest, g is a link function, G is a SNP dose, and $\mathbf{Z}$ is a vector of adjustment variables; the coefficient of interest is α_G. The coefficients in this model can be estimated by linear regression for quantitative outcomes (g=identity) and logistic regression for binary outcomes (g=logit). To investigate gene-environment interactions with single SNPs, comparable models are used [12]:

$$g(E[Y_i])=\alpha_0+\beta_E E_i+\beta_G G_i+\beta_{G:E} E_i G_i+\boldsymbol{\gamma}\boldsymbol{Z}_{\mathbf{i}} \quad (2)$$

where E is an environmental exposure, and the coefficient of interest is $\beta_{G:E}$.

These cross-sectional GWAS models have low power to detect novel associations when genetic variants and/or environmental exposures are rare [3, 13]. One way to increase power, as we will demonstrate in Section 2.4, is to use the repeated outcome and exposure measurements that are often available in cohort data. Indexing the multiple visits by measurement time t, the cross-sectional models can be revised as follows:

$$g(E[Y_{it}])=\alpha_0+\alpha_G G_i+\boldsymbol{\gamma}\boldsymbol{Z}_{\mathbf{it}} \quad (3)$$

$$g(E[Y_{it}])=\beta_0+\beta_E E_{it}+\beta_G G_i+\beta_{G:E} E_{it} G_i+\boldsymbol{\gamma}\boldsymbol{Z}_{\mathbf{it}} \quad (4)$$

where α_G and $\beta_{G:E}$ remain the coefficients of interest. In the genetics literature, the mean models in equations (1) and (3) are referred to as *main effects* models and the ones in equations (2) and (4) are *interaction* models.

2.2. Generalized Estimating Equations

When repeated measures of outcome and exposure are used, methods to estimate coefficients of interest must allow for the correlated nature of the data. Common methods that allow for correlation are generalized estimating equations (GEE) and mixed effects models (MEM) [1]. Although both options may be relevant in GWAS, in this manuscript we focus on GEE.

GEE is a semi-parametric method that requires assumptions about the form of the mean of the outcome distribution, conditional on covariates, but does not require assumptions about the full conditional distribution of the outcome [14, 15]. Correct specification of the covariance of repeated measurements within each person is not required for asymptotically-valid inference, but instead, a “working” covariance matrix is assumed, which determines how data points are weighted in the resulting inference. Robust variance estimators are used to obtain valid inference. Although specification of the correct correlation matrix is not generally required, careful consideration must be given to the choice of correlation when covariates vary over time [16]. Specifically, if the marginal expectation of the outcome at time t conditional on covariate values at time t is not equal to the marginal expectation of the outcome at time *t* conditional on covariate values observed at all times, then a working independence correlation matrix should be assumed for validity of estimates. If, however, covariates do not vary over time, or the specified assumption is satisfied, then a correlation matrix that more closely reflects the true underlying correlation will provide more efficient parameter estimates [14].

In general, parameters estimated via GEE have population-averaged (marginal) interpretations, whereas those estimated via MEM are participant-specific (conditional) summaries. In the special case of quantitative outcomes, where the identity link function is collapsible, MEM parameters have population-averaged interpretations as well; however, the same is not true for outcomes that require non-collapsible link functions [17]. For example, with binary outcome data recorded longitudinally and using the logistic link function, the parameter α_G is the log odds ratio comparing individuals with one additional versus one fewer copy of the minor allele. A comparable parameter obtained using MEM would be the log odds ratio comparing individuals with a shared (unobserved) factor who have one additional copy of the minor allele to those with the same factor and one fewer copy. When the goal of a large-scale genetic analysis is to characterize population-level associations, then GEE is more appropriate than MEM for many non-quantitative outcomes.

However, compared to MEM, GEE has the disadvantage of requiring stronger assumptions about missing data [1]. The key covariate, genotype, is assumed to be constant over time, with missing data minimized via imputation to a common reference panel. For MEM to be valid, the phenotypic data need to be at worst missing at random (MAR), conditional on modeled covariates, and with a correctly specified covariance model. For GEE to be valid, phenotypic data need to be missing completely at random (MCAR), conditional on modeled covariates. The required assumptions relate to the interpretation of parameters as marginal versus conditional; GEE is not as robust to differential death or dropout because the population being averaged over at later times could be different. However, the improved robustness to missing phenotypic data in MEM with a correct covariance model comes at the price of treating death and missingness as the same [18].

Methods exist for weakening the MCAR assumption underlying GEE [19]; however, they are not straightforward to implement consistently across studies and meta-analyze, which is the mechanism of analysis in many genome-wide investigations [10]. Further, the problem of missingness may not be a substantial one in GWAS analyses because the effects of individual genetic variants are not generally strong enough to determine who enters the study or who drops out from phenotype measurement over time.

2.3. Computation

One challenge associated with implementing longitudinal methods in the context of GWAS is the computational complexity required to fit them. On a typical desktop computer, using data on several thousand participants, fitting a single GEE or MEM model takes around a second; however, fitting millions of them can require hundreds of hours of computing time. Therefore it is crucial to develop and utilize algorithms that minimize computational burden. Substantial efforts have been made to minimize computational time for MEM [20, 21, 22, 23], and use of a one-step estimator [24] has recently been proposed to minimize computational time for GEE [25]. Even without using the one-step approximation, custom code for fitting GEE models in R, available in the `boss` package [25], can result in nearly 10-fold decreases in computing time, compared with standard implementations.

2.4. Power: Cross-sectional versus Longitudinal

Longitudinal analyses take fuller advantage of the data that have been collected in longitudinal cohort studies, using the additional within-person information to achieve increases in statistical power to detect associations. The biggest gains in power occur when there is more within-person variability in outcome in genetic analyses of main effects (equations (1) and (3)) and more within-person variability in environmental exposure in analyses of gene-environment interactions (equations (2) and (4)). Such increases in power are particularly valuable in the context of genome-wide analyses that must use low thresholds for statistical significance in order to account for the large number of comparisons that are performed.

Increases in power with use of longitudinal data are well-known, and have been illustrated previously in the context of the common significance threshold of 0.05 [1]. However, relative power may be different at lower thresholds than it is at the common threshold of 0.05 [26], so we conducted simulations to confirm the expected gains in power from using longitudinal data in genome-wide analyses. Results for detecting gene-drug interactions were included in a recent manuscript describing cross-sectional analyses of the QT interval done by the CHARGE pharmacogenetics group [2]. Results for detecting main effects of rarer variants are included in Supplemental Figure 1. The outcome was assumed to be normal, with a two-sided, per-SNP $\alpha = 5.0 \times 10^{-8}$ and MAF=0.01. The effect to be detected varied in terms of standard deviations of the outcome. An attrition rate of 5% per visit was assumed, plus random missingness of 5% of remaining measurements. To evaluate the power that could be gained by incorporating repeated measures over time, the simulations incorporated up to 2–6 measurements of QT duration for each of 20,000 participants, and within-person correlation in QT varied from 0.2 to 0.8. Both exchangeable and autoregressive (AR1) correlation structures were examined. Potential gains in power can be larger with AR1 correlation because there is more within-person variation across measurements. Linear models with robust standard errors were used for cross-sectional analyses, and generalized estimating equations (GEE) with independence working correlation and sandwich standard error estimates were used for longitudinal analyses. For example, for such rarer main effects, with n=20,000 and 80% power, the detectable effect is nearly 0.35 standard deviations using cross-sectional data, but decreases to about 0.25 standard deviations using 4 repeated measures with exchangeable correlation of 0.5 (Supplemental Figure 1). Comparable decreases in detectable effects exist for gene-environment interactions. Given the sample size of 20,000 and MAF of 0.01, each set of simulated cross-sectional data includes approximately 400 observations with a minor allele, a sufficient number for asymptotic results with robust variance estimates to be valid. In the following section, we discuss scenarios in which asymptotic results may not be valid.

2.5. Small effective sample sizes: options for valid inference

In scenarios with rare variants and/or rare environmental exposures, where the increased power derived from longitudinal analyses is most needed, the asymptotics of robust sandwich variance estimators used in GEE are not sufficiently accurate due to small effective sample sizes [27]. When the ratio $\frac{\hat{\beta}}{\sqrt{\widehat{\mathrm{Var}}[\hat{\beta}]}}$ is assumed to have a normal reference distribution, type I error can be inflated relative to nominal values [27]; our simulations, described below, illustrate that this inflation of type I error can be even more dramatic at the low significance levels used in GWAS than it is at the more common significance level of 0.05.

Several options for improving the accuracy of the significance statements made with GEE, in the context of small effective sample sizes, have been proposed in the literature. These options include (i) using an alternate reference distribution for the test statistic that incorporates the variability in standard error estimates [27], (ii) using an alternate standard error estimate [28, 29, 30], and (iii) using a score test instead of a Wald test [31]. We assess all three of these methods in simulations, but focus on the first, where we assume that the ratio of the coefficient estimate to its standard error follows a t reference distribution, which has heavier tails than the normal distribution [32]. Use of a t reference distribution requires an estimate of degrees of freedom; one option, based on Satterthwaite's approximation [33], is discussed by Pan and Wall [27]. Instead of ignoring the variability in standard error estimates, as is done when using a normal reference distribution, the degrees of freedom estimate for the t distribution accounts for this variability. Specifically, using moment-matching arguments, the degrees of freedom equal

$$2\frac{E[\mathrm{Var}(\hat{\beta})]^2}{\mathrm{Var}[\mathrm{Var}(\hat{\beta})]}.$$

Pan and Wall assume that the variance estimate $\widehat{\mathrm{Var}}[\hat{\beta}]$ is unbiased, but now incorporate an estimate of its variability in the denominator of the degrees of freedom estimate [27]. The `boss` package in R includes the option to calculate Pan and Wall's estimate of degrees of freedom. It minimizes the computational time by implementing C code that is called from within the relevant R function.

To illustrate the inflated significance of p-values calculated using a normal reference distribution, as well as the performance of methods to correct these p-values, we simulated data for n=1000 people, each with 4 exchangeable outcome measurements, within-person correlation of 0.5, and no associations between a single SNP and outcomes. For longitudinal main effects models (equation (3)), we used MAF=0.01; for longitudinal gene-drug interaction models (equation (4)), we used MAF=0.05 and proportion of participants using the drug, independent of genotype, at any given measurement time = 0.07.We then performed GEE analysis and calculated five p-values: (i) using sandwich standard error estimates and the standard normal reference distribution, (ii) using sandwich standard error estimates and a t reference distribution with Satterthwaite estimates of degrees of freedom

[27], (iii) using Mancl and DeRouen's alternate standard error estimates [28] and a t reference distribution with Satterthwaite estimates of degrees of freedom as described byWang and Long [30], (iv) using Wang and Long's alternate standard error estimates and a normal reference distribution [30], and (v) using a robust score test [31].We repeated the process for a million simulated datasets, and generated quantile-quantile plots of the resulting p-values on the −log10 scale (Figure 1).

Using sandwich standard error estimates with a normal reference distribution, as the p-values decrease, there is substantial deviation in the observed p-values compared to what is expected under the null. For the interaction models, the point estimates are unbiased for the true value of zero (Supplemental Figure 4a), and the standard error estimates are also essentially unbiased for the true standard error of 0.2 (Supplemental Figure 4b). However, the variance of the standard error estimates is 0.03, which is non-negligible compared to the variability of the coefficient estimates. In standard asymptotic approximations the smaller variance is treated as zero, and this results in p-values that over-state the true significance.

When the model is correctly specified, using either a t reference distribution or alternate standard error estimates largely corrects the problem. The score test over-corrects, giving conservative p-values. Mancl and DeRouen's alternate standard error estimates give similar performance to sandwich variance estimates, and require additional computing time, so we focus on sandwich estimates for ease of implementation. Wang and Long's alternate standard error estimates require stronger assumptions for validity: specifically that the variance of the outcome, conditional on covariates, is correctly specified and that there is a common correlation structure across all participants, or at least across all participants with the same number of observations. In this respect, the standard error estimates obtained with Wang and Long's methods are closer to the model-based estimates, which no longer control type I error when models are mis-specified. We illustrate this drawback by simulating data from interaction models where the variance of outcomes is different among those in the exposed group versus the unexposed group; in such a setting, type I error is not controlled with Wang and Long's method, but it is controlled with the other two robust methods (Figure 2). Given that the stronger assumptions are not likely to be satisfied in our intended applications, we choose to focus on modification of the reference distribution alone.

2.6. Small effective sample sizes: conditions for valid inference

Additional simulations varying MAF and probability of drug use illustrate that, when the product of sample size and MAF is small, even the small-sample correction using the t reference distribution on which we focus in Section 2.5 leaves substantial p-value inflation in GEE analyses (Supplemental Figures 2 and 3) – reflecting the unsatisfactory behavior of the asymptotic approximations being used. Such behavior occurs in both cross-sectional and longitudinal analyses when robust sandwich standard error estimates are used. One way to characterize this behavior is in terms of the number of independent observations in the smallest group being compared, or in other words as a simple estimate of the degrees of freedom the data provide for inference, denoted 'approxdf'. This characterization comes from analogy to a t test with unequal variances across groups, where if the smallest group is substantially smaller than the others, the Satterthwaite approximation of degrees of freedom approaches the size of the smallest group. For main effects models (equations (1) and (3)), this smallest group is the number of independent observations in people who have the minor allele for a SNP of interest, which can be approximated as follows: approxdf = 2 × MAF × Nindep, where Nindep = the estimated number of independent observations. For gene-environment interaction models (equations (2) and (4)) with binary exposure, approxdf = 2×MAF×Nexposed, where Nexposed = the estimated number of independent observations in the smaller of the exposed or nonexposed groups [34].

With longitudinal data, one way to compute Nindep is to sum over the estimated number of independent observations per person [35], that is

$$\text{Nindep}=\sum_i \frac{n_i}{1+(n_i-1)\hat{\rho}},$$

where n_i is the number of observations for participant i and $\hat{\rho}$ is an estimate of the pairwise visit-to-visit correlation within participants from a GEE-exchangeable model that does not contain genetic data. For gene-environment interactions with binary exposure, assuming that the proportion of people who are exposed is less than one-half,

$$\text{Nexposed}=\sum_i \frac{n_i}{1+(n_i-1)\hat{\rho}} \frac{\#\{E_{it}=1\}}{n_i},$$

where $\#\{E_{it} = 1\}$ is the number of observations in which participant i is exposed.

Genetic data used for GWAS are typically imputed to a common reference panel to facilitate pooling of information across studies. For imputed genetic data, following Rubin’s characterization of the fraction of missing information [36], the approxdf can be obtained by multiplying the value obtained for non-imputed data by a SNP-specific measure of imputation quality, specifically the ratio of the observed variance of imputed allele counts to their expectation based on estimated allele frequencies [10].

Based on Supplemental Figures 2 and 3, in genome-wide work, we recommend requiring that SNPs have within-study approxdf greater than 10 in order to enter across-study meta-analyses. Although study-specific contributions still have some inflation at approxdf=10, we have found that in the context of our CHARGE multi-study meta-analysis this threshold provides a good balance between over- and under- filtering. The core studies in CHARGE each include at least several thousand participants; a meta-analysis that included only smaller studies might need to use a more stringent filter to avoid spurious results at the meta-analytic level. Despite the need to ignore results below a particular level of approxdf, the use of Satterthwaite estimates of degrees of freedom provides advantages over standard approaches that filter on MAF or oevar alone (not their product, or prevalence of drug exposures). The Satterthwaite approach provides more accurate p-values than standard GEE methods and thereby reduces the number of variants that would not be considered, compared to the harsh filtering required when considering MAF or oevar alone to ensure that standard GEE methods are correctly calibrated.

2.7. Meta-analysis

To achieve useful levels of power, most genome-wide studies require sample sizes that are larger than exist in any single study, but individual-level data cannot typically be combined across studies due to both privacy concerns and incompatibility of study designs or data elements. Therefore in GWAS individual studies generally analyze their own data, then share summary information for meta-analysis across studies [10]. Meta-analysis typically uses fixed effects inverse-variance-weighted methods, which are known to be statistically efficient in common situations [37]. Coefficient and variance estimates from GEE analyses can also be meta-analyzed in this way. However, when small effective sample sizes in individual studies cause concern about inflation of significance of study-specific association results, the inflation can be propagated through standard meta-analysis of results, even with genomic control correction. Meta-analysis methods that mitigate this inflation include 1) comparing z-statistics computed from inverse-variance-weighted meta-analysis to a t distribution with degrees of freedom equal to the sum of the individual-cohort degrees of freedom and 2) computing individual-cohort p-values using the t reference distribution, then meta-analyzing them using a weighted z-statistic, with weights equal to the product of the SNP imputation quality and the study-specific estimated number of independent observations (or the number exposed to the drug, for gene-environment interactions). One drawback to the latter approach is the lack of a meta-analytic estimate of the coefficient of interest.

To evaluate the three options for meta-analysis (fixed effects inverse-variance-weighted, plus the two alternatives described above), we simulated null data in three different cohorts, of size 1000, 1000, and 2000 individuals, then metaanalyzed results. We used the gene-environment interaction set-up from Section 2.5, and again ran one million iterations. Figure 3 displays cohort-specific and meta-analytic qq-plots on the −log10 scale for p-values less than 0.001. Although combining data across studies ameliorates the inflation seen in individual cohorts, meta-analytic p-values based on a normal reference distribution still have inflated type I error. Depending on the specific scenario of interest, inverse-variance-weighted meta-analyses using a t-distribution with degrees of freedom equal to the sum of individual-cohort degrees of freedom may be valid; however, weighted z-value based meta-analysis of p-values using t reference distributions at the cohort level appears to be the safest option. Results were similar for simulations of main effects analyses.

3. Application

Pharmacogenetics refers to genetic differences that lead to differences in individuals' responses to drugs. A number of drugs are known to prolong the QT interval [38, 39], and QT prolongation is a risk factor for ventricular tachyarrythmias such as Torsades de pointes and sudden cardiac death [40, 41]. As QT interval is heritable [42], and there are interindividual differences in QT prolongation among users of drugs that prolong the QT interval [39], it seems plausible that genetics plays a role in drug-induced QT prolongation. Thus one goal of the CHARGE consortium's pharmacogenetics group is to evaluate gene-drug interactions on QT interval, for variants across the entire genome. Drugs of interest include thiazide diuretics, sulfonylureas, tri and tetracyclic antidepressants (TCAs), and definite or possible QT prolonging drugs, as identified by the University of Arizona Center for Education and Research on Therapeutics (UAZ CERT) [43].

The CHARGE consortium consists of cohort studies focused on cardiovascular outcomes, with detailed measurements on a range of outcomes and exposures, including genome-wide genetic data [4]. Many of the cohort studies have repeated measurements on the same participants, enabling gains in power from use of longitudinal data, gains which are particularly important in the context of gene-environment interaction. To illustrate the methods described in this manuscript, we present analyses of data from the Cardiovascular Health Study, a cohort of adults 65 years and older, enrolled in 1989–1990 at sites across the United States, and followed through annual visits for ten years, then phone contacts thereafter [5]. QT interval was electronically measured from automatically processed, annual, digital ECG recordings. Prescription drug use was determined concurrently. Genotyping was done using the Illumina 370CNV BeadChip system on all CHS participants who were free of cardiovascular disease and consented to genetic testing; data were imputed at each of the 2.5 million autosomal HapMap SNPs.

Up to 3055 participants, with an average of 7 observations per person, are included in analyses of gene-drug interactions on QT interval in those of European descent in CHS. The following mean model was used:

$$E[QT_{it}]=\beta_0+\beta_E\mathrm{drug}_{it}+\beta_G\mathrm{SNP}_i+\beta_{G:E}\mathrm{drug}_{it}\mathrm{SNP}_i+\boldsymbol{Z}'_{\mathbf{it}}\boldsymbol{\gamma}$$

where covariates **Z** include age, gender, RR interval, ancestry, and use of UAZ CERT QT-prolonging drugs (for the three non-UAZ CERT drug groups). The coefficient of interest is $\beta_{G:E}$, and it is estimated using GEE with working independence correlation. Drug exposure is measured separately at each visit, so that some participants may be exposed at one time, but not at other times. Given the possibility that previous drug exposure may influence future outcomes, and that these outcomes may influence subsequent drug exposures, working independence is chosen to ensure validity of parameter estimates [16]. GEE parameter estimates, with their marginal interpretations, may be biased if QT measurements are differentially missing across drug exposure and genotype. However, we do not expect genetic effects to be large enough for this missingness to substantially impact the results presented for the interaction term.

Figure 4 illustrates the qq-plots for interaction p-values, on the −log10 scale, in GEE analyses of the four drugs of interest. Among these four drugs, exposure to thiazide diuretics is most common (18%), followed by UAZ CERT QT-prolonging drugs (5.4%), sulfonylureas (4.9%), and TCAs (3.5%). As described in Section 2.5, when using a standard Normal reference distribution, substantial inflation of type I error occurs at low significance levels when environmental exposure is rare. Such inflation is not evident in analyses of the more commonly-used thiazides, but is evident in analyses of other drug classes. To better approximate the statistical significance of the results, we re-calculated the p-values using the methods described in Section 2.5 (Figure 4). Given the lack of inflation in the thiazides analyses, any of the methods is appropriate in that context (Figure 4a). In the analyses of

UAZ CERT drugs, Wang and Long's method shows inflation of p-values similar to that seen in the simulations with heteroscedasticity, whereas the other methods perform comparably to each other (Figure 4b). In the analyses of sulfonylureas and TCAs, model misspecification appears not to be an issue, as Wang and Long's method performs best, followed by Mancl and DeRouen's variance estimates and typical sandwich variance estimates with a t reference distribution (Figure 4c and 4d). For the rarest drugs, filtering out SNPs with approximate degrees of freedom equal to 10 or less alleviates residual inflation with typical sandwich variance estimates (Figure 4, light gray). More SNPs are filtered due to low approximate degrees of freedom in analyses of rarer drugs (Supplemental Figure 5).

To illustrate a meta-analysis of several longitudinal GWAS, we combined results from CHS, ARIC, and RS for the UAZ CERT drug exposure. Results from both the original RS (RS1) and an additional cohort recruited ten years later (RS2) are included. Nexposed for CHS was 378, for ARIC was 550, for RS1 was 303, and for RS2 was 105. Study-specific qq-plots of p-values on the −log10 scale for RS1, RS2, and ARIC are shown in Supplemental Figure 6. As in CHS, Wang and Long's method gives inflated p-values; in ARIC and RS1, other methods perform comparably, while in RS2, with its lower Nexposed, filtering is needed to remove inflation. Study-specific genomic control correction was applied to all meta-analyses, which use typical sandwich variance estimates. When all SNPs were included, both a standard fixed effects inverse-variance-weighted meta-analysis and a weighted z-based meta-analysis had inflated type I error (Figure 5). When study-specific results were filtered to include only those SNPs with approximate degrees of freedom greater than 10, the inverse-variance-weighted meta-analysis still had inflated type I error, while the weighted z-based meta-analysis that used a t reference distribution to calculate p-values did not (Figure 5). The distributions of approximate degrees of freedom are displayed in Supplemental Figure 7; part (e) of the figure illustrates that the summation of degrees of freedom across studies in meta-analysis will improve the asymptotic accuracy. Although no SNPs were found to have genome-wide significant interactions in this analysis, additional cohorts will contribute to final analyses in CHARGE.

4. Discussion

Longitudinal data can be used to increase power to detect associations in GWAS, especially in the low-power settings of rarer variants and gene-environment interactions. Methods tailored to longitudinal data, such as GEE or MEM, must be used. In contexts where GEE is appropriate, the R `boss` package requires reasonably low computational time and implements a correction for small effective sample sizes that modifies the reference distribution for the test statistic of interest. Through application of such methods to data from the CHARGE consortium pharmacogenetics group, we have shown that using the small-sample correction in individual cohorts, then conducting weighted Z-based meta-analysis with modest filtering, yields well-calibrated genome-wide results.

As discussed in Section 2.5, other options for small-sample corrections include modifying the standard error estimates and implementing score tests. It is well-known that the score test is conservative for small numbers of clusters [31], and we confirm that this property holds in the context of small effective sample sizes. The modification proposed by Guo et al. [31]

does not apply in the contexts that we consider because the total number of clusters is not small, but the number of clusters with a minor allele (in main effects analysis) or with an exposed measurement and a minor allele (in interaction analyses) is small. Further work could be done to devise a correction in this context, but such work is outside the scope of this paper. The different Wald tests that modify the standard error estimates and/or the reference distribution have similar performance in the small-sample contexts that we consider. Choosing from among these tests requires balancing the required computational time with the incremental benefit in performance, when comparing typical sandwich estimates with their modifications; as well as assessing the plausibility of the additional assumptions, particularly when considering Wang and Long’s modification. Due to the potential for heteroscedasticity in outcome variance and lack of common correlation structure in unbalanced cohort data, we generally recommend typical sandwich variance estimates, in conjunction with modification of the reference distribution, in this context. However, in situations where Wang and Long’s assumptions are likely to be satisfied, there is potential for improved performance with their use.

One potential limitation to the use of GEE in the context of gene-environment interactions is the possibility of selection bias when environmental exposure varies over time. For example, when the exposure is drug use, there may be differences in factors associated with initiation and cessation of treatment, both for a particular therapeutic, and for specific agents within therapeutic classes. Our choice of independence working correlation should decrease the impact of such selection bias; however, more sophisticated methods such as inverse-probability weighting [44] would do more to alleviate concerns. As with similar methods that can weaken missingness assumptions, such methods are challenging to implement in the context of GWAS, and deserve further consideration.

We have not discussed use of MEM, though it is a reasonable alternative approach, particularly in the case of quantitative phenotypes where MEM parameter estimates have a marginal interpretation comparable to that of GEE estimates. MEM can provide more robustness to bias caused by some types of missing data, but they also rely more heavily on modeling assumptions [45]. If, for example, random effects are correlated with covariates such as ancestry, then bias in coefficient estimates could be substantial, leading to substantially misleading inference.

Another natural extension is to family data; methods for analyzing multi-level longitudinal family data include family-based association tests [46] and Markov chain Monte Carlo (MCMC) implementations of generalized linear mixed models (GLMMs) [47]. However, neither of these approaches is well-suited to the context of genome-wide analysis with small effective sample sizes. Therefore further work is needed to determine the best methods for analyzing longitudinal data collected on related individuals in a genome-wide manner.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank the reviewers and associate editor for their valuable comments and suggestions that improved the quality of our manuscript. This work was supported by US NIH R01 HL103612 (Dr. Psaty). Study-specific support included the following:

ARIC: The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

CHS: This Cardiovascular Health Study research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL105756 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the National Institute on Aging (NIA). A full list of CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm.

DNA handling and genotyping at Cedars-Sinai Medical Center was supported in part by the National Center for Research Resources, grant UL1RR033176, and is now at the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124; in addition to the National Institute of Diabetes and Digestive and Kidney Disease grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

RS:The GWA database of the Rotterdam Study was funded through the Netherlands Organisation of Scientific Research NWO (nr. 175.010.2005.011). The Rotterdam Study is supported by the Erasmus Medical Center and Erasmus University, Rotterdam; the Netherlands Organization for Scientific Research (NWO), the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. Specifically, work on this project was supported by ZonMW Priority Medicines Elderly Program (nr. 113101002).

References

1. Diggle, P.; Heagerty, P.; Liang, KY.; Zeger, S. Analysis of Longitudinal Data. 2nd edn.. Oxford: Oxford University Press; 2002.
2. Avery C, Sitlani C, Arking D, Arnett D, Bis J, Boerwinkle E, Buckley B, Chen I, de Craen A, Eijgelsheim M, et al. Drug-gene interactions and the search for missing heritability: a cross-sectional pharmacogenomics study of the QT interval. The Pharmacogenomics Journal. 2013
3. Asimit J, Zeggini E. Rare Variant Association Analysis Methods for Complex Traits. Ann Rev Genet. 2010; 44:293–308. [PubMed: 21047260]
4. Psaty B, O'Donnell C, Gudnason V, Lunetta K, Folsom A, Rotter J, Uitterlinden A, Harris T, Witteman J, Boerwinkle E. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from five cohorts. Circ Cardiovasc Genet. 2009; 2:73–80. [PubMed: 20031568]
5. Fried L, Borhani N, Enright P, Furberg C, Gardin J, Kronmal R, Kuller L, Manolio T, Mittelmark M, Newman A, et al. The Cardiovascular Health Study: design and rationale. Ann Epidemiol. 1991; 1(3):263–276. [PubMed: 1669507]
6. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol. 1989; 129:687–702. [PubMed: 2646917]
7. Hofman A, Grobbee D, de Jong P, van den Ouweland F. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. Eur J Epidemiol. 1991; 20:403–422. [PubMed: 1833235]
8. Hofman A, van Duijn C, Franco O, Ikram M, Janssen H, Klaver C, Kuipers E, Nijsten T, Stricker B, Tiemeier H, et al. The Rotterdam Study: 2012 objectives and design update. Eur J Epidemiol. 2011; 26(8):657–686. [PubMed: 21877163]

9. Burton P, Clayton D, Cardon L, Craddock N, Deloukas P, Duncanson A, Kwiatkowski D, McCarthy M, Ouwehand W, Samani N, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447(7145):661–678. [PubMed: 17554300]

10. de Bakker P, Ferreira M, Jia X, Neale B, Raychaudhuri S, Voight B. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008; 17:R122–R128. [PubMed: 18852200]

11. McCullagh, P.; Nelder, J. Generalized Linear Models. 2nd edn.. Boca Raton: Chapman and Hall / CRC; 1989.

12. Thomas D. Methods for investigating gene-environment interactions in candidate pathway and genome-wide association studies. Annu Rev Public Health. 2010; 31:21–36. [PubMed: 20070199]

13. Khoury M, Wacholder S. From Genome-Wide Association Studies to Gene-Environment-Wide Interaction Studies Challenges and Opportunities. Am J Epidemiol. 2009; 169:227–230. [PubMed: 19022826]

14. Liang KY, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.

15. Zeger S, Liang KY. Longitudinal Data Analysis for Discrete and Continuous Outcomes. Biometrics. 1986; 42:121–130. [PubMed: 3719049]

16. Pepe M, Anderson G. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. Commun Statist - Simula. 1994; 23(4):939–951.

17. Zeger S, Liang KY, Albert P. Models for Longitudinal Data: A Generalized Estimating Equation Approach. Biometrics. 1988; 44:1049–1060. [PubMed: 3233245]

18. Kurland B, Heagerty P. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. Biostatistics. 2005; 6(2):241–258. [PubMed: 15772103]

19. Robins J, Rotnitzky A, Zhao L. Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. JASA. 1995; 90(429):106–121.

20. Kang H, Sul J, Service S, Zaitlen N, y Kong S, Freimer N, Sabatti C, Eskin E. Variance component model to account for sample structure in genomewide association studies. Nature Genetics. 2010; 42:348–354. [PubMed: 20208533]

21. Zhang Z, Ersoz E, Lai CQ, Todhunter R, Tiwari H, Gore M, Bradbury P, Yu J, Arnett D, Ordovas J, et al. Mixed linear model approach adapted for genome-wide association studies. Nature Genetics. 2010; 42:355–360. [PubMed: 20208535]

22. Meyer K, Tier B. SNP Snappy: A Strategy for Fast Genome Wide Association Studies Fitting a Full Mixed Model. Genetics. 2011; 190:275–277. [PubMed: 22021386]

23. Pirinen M, Donnelly P, Spencer C. Efficient Computation with a Linear Mixed Model on Large-Scale Data Sets with Applications to Genetic Studies. Annals of Applied Statistics. 2012 in press.

24. Lipsitz S, Fitzmaurice G, Orav E, Laird N. Performance of Generalized Estimating Equations in Practical Situations. Biometrics. 1994; 50:270–278. [PubMed: 8086610]

25. Voorman A, Rice K, Lumley T. Fast computation for genome-wide association studies using boosted one-step statistics. Bioinformatics. 2012; 28(14):1818–1822. [PubMed: 22592383]

26. Morris N, Elston R. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. A Note on Comparing the Power of Test Statistics at Low Significance Levels. 2011; 65(3):164–166.

27. Pan W, Wall M. Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. Statist Med. 2002; 21:1429–1441.

28. Mancl L, DeRouen T. A Covariance Estimator for GEE with Improved Small-Sample Properties. Biometrics. 2001; 57:126–134. [PubMed: 11252587]

29. Pan W. On the robust variance estimator in generalised estimating equations. Biometrika. 2001; 88:901–906.

30. Wang M, Long Q. Modified robust variance estimator for generalized estimating equations with improved small-sample performance. Statist Med. 2011; 30:1278–1291.

31. Guo X, Pan W, Connett J, Hannan P, French S. Small-sample performance of the robust score test and its modifications in generalized estimating equations. Statist Med. 2005; 24:3479–3495.

32. Brazzale, A.; Davison, A.; Reid, N. Applied Asymptotics: Case Studies in Small-Sample Statistics. New York: Cambridge University Press; 2007.

33. Satterthwaite F. An Approximate Distribution of Estimates of Variance Components. Biometrics Bulletin. 1946; 2(6):110–114. [PubMed: 20287815]

34. Good I. What are Degrees of Freedom? The American Statistician. 1973; 27(5):227–228.

35. Hanley J, Negassa A, Edwardes M, Forrester J. Statistical Analysis of Correlated Data Using Generalized Estimating Equations: An Orientation. Am J Epidemiol. 2003; 157:364–375. [PubMed: 12578807]

36. Horton N, Lipsitz S. Multiple imputation in practice: comparison of software packages for regression models with missing variables. The American Statistician. 2001; 55(3):244–254.

37. Lin D, Zeng D. On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. Biometrika. 2010; 97(2):321–332. [PubMed: 23049122]

38. Yap Y, Camm A. Drug induced QT prolongation and torsades de pointes. Heart. 2003; 89(11): 1363–1372. [PubMed: 14594906]

39. Roden D. Drug induced prolongation of the QT interval. N Engl J Med. 2004; 350:1013–1022. [PubMed: 14999113]

40. Schouten E, Dekker J, Meppelink P, Kok F, Vandenbroucke J, Pool J. QT interval prolongation predicts cardiovascular mortality in an apparently healthy population. Circulation. 1991; 84(4): 1516–1523. [PubMed: 1914093]

41. Straus S, Kors J, Bruin MD, van der Hooft C, Hofman A, Heeringa J, Deckers J, Kingma J, Sturkenboom M, Stricker B, et al. Prolonged QTc interval and risk of sudden cardiac death in a population of older adults. J Am Coll Cardiol. 2006; 47(2):362–367. [PubMed: 16412861]

42. Newton-Cheh C, Eijgelsheim M, Rice K, de Bakker P, Yin X, Estrada K, Bis J, Marciante K, Rivadeneira F, Noseworthy P, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nature Genetics. 2009; 41(4):399–406. [PubMed: 19305408]

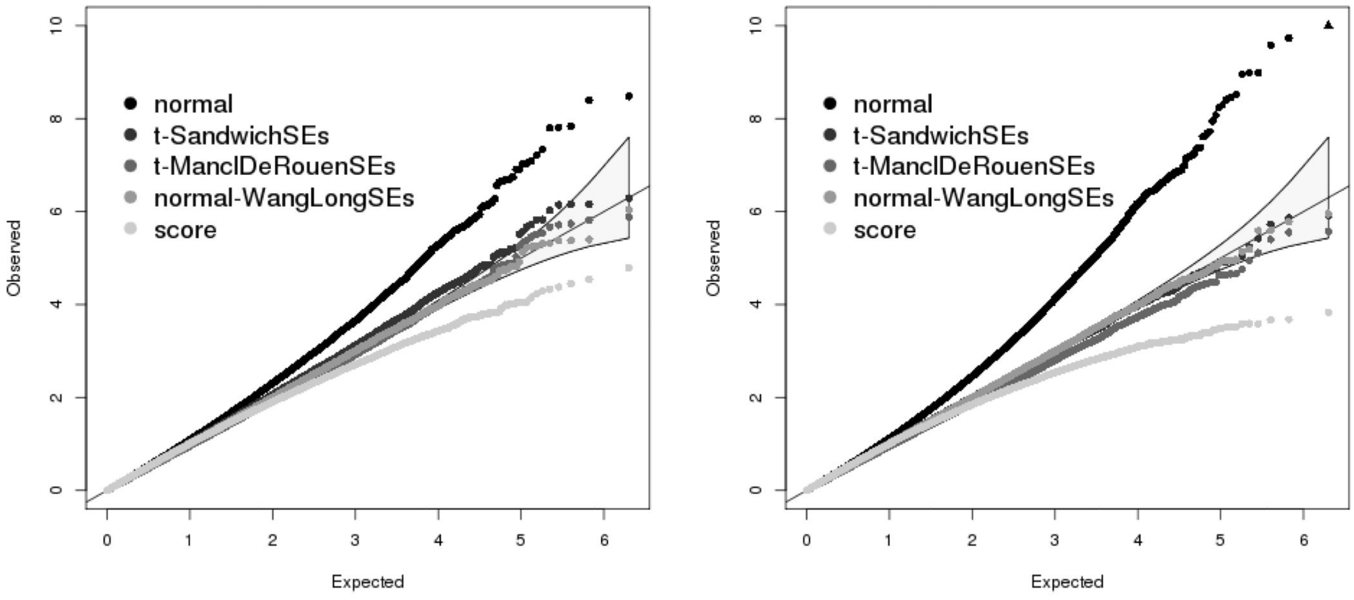
43. Drugs that Prolong the QT Interval. AZ, USA: Credible Meds – AZCERT; 2011. Http://www.azcert.org/.

44. Robins J, Hernan M, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. Epi. 2000; 11(5):550–560.

45. Heagerty P, Kurland B. Misspecified maximum likelihood estimates and generalized linear mixed models. Biometrika. 2001; 88:973–985.

46. Ding X, Lange C, Xu X, Laird N. New Powerful Approaches for Family-based Association Tests with Longitudinal Measurements. Ann Huml Genet. 2009; 73:74–83.

47. Burton P, Scurrah K, Tobin M, Palmer L. Covariance components models for longitudinal family data. Int J Epidemiol. 2005; 34:1063–1077. [PubMed: 15831561]

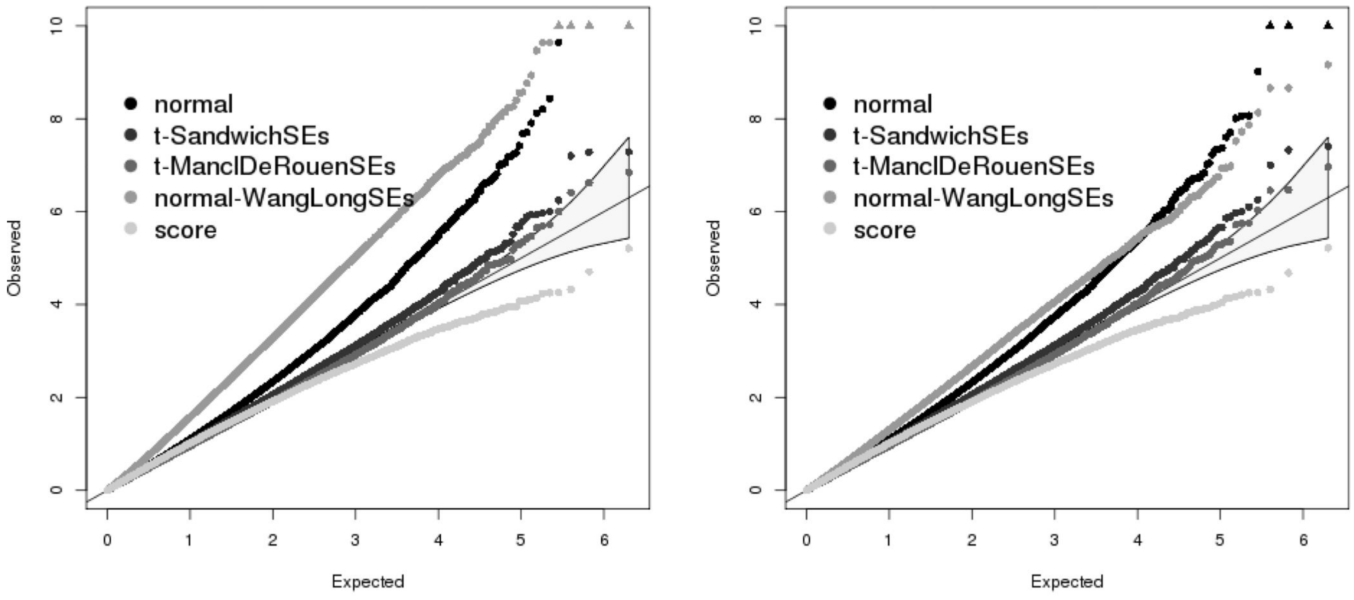


(**a**) Gene-environment interaction models.

(**b**) Main effect models.

Figure 1.

Data from 1 million iterations were simulated under the null model, for 1000 individuals, each with 4 exchangeable outcome measurements, and within-person correlation of 0.5. For the analysis of gene-environment interaction models (a), MAF was 0.05 and proportion exposed at any given visit was 0.07; for the analysis of main effects models (b), MAF was 0.01. P-values were estimated via GEE with an independence working correlation matrix, in the following contexts: (i) using sandwich standard error estimates and the standard normal reference distribution, (ii) using sandwich standard error estimates and a t reference distribution with Satterthwaite estimates of degrees of freedom, (iii) using Mancl and DeRouen's alternate standard error estimates and a t reference distribution, (iv) Wang and Long's alternate standard error estimates and a normal reference distribution, and (v) using a score test. Plots are quantile-quantile plots of −log10(pvalues), with the cone indicating the 95% prediction interval for the ordered −log10(p-values), when p-values are truly uniformly distributed between zero and one.

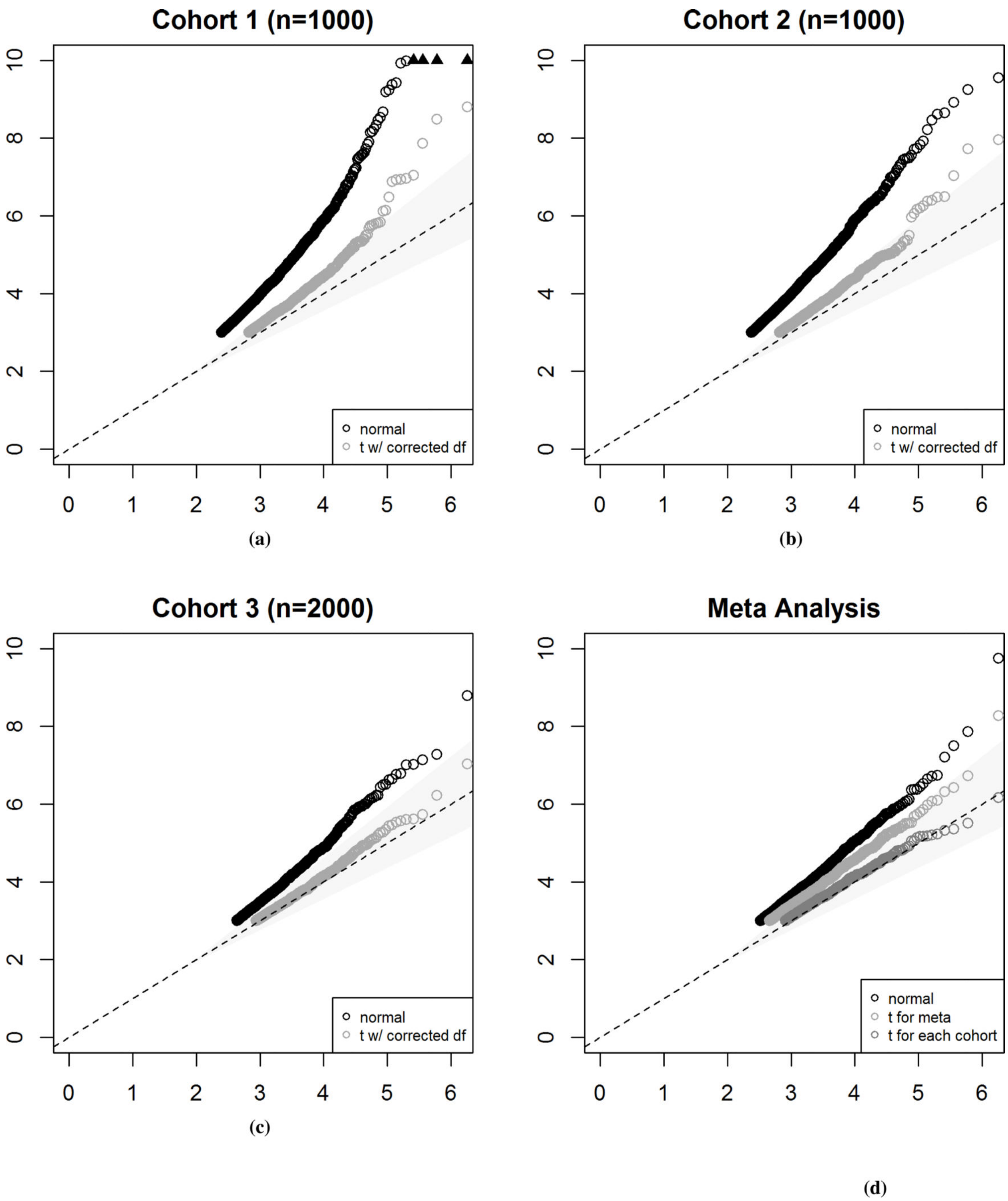


(a) Variance 100% higher in exposed group.

(b) Variance 50% higher in exposed group.

Figure 2.

Gene-environment interaction data from 1 million iterations were simulated under the null model, for 1000 individuals, each with 4 exchangeable outcome measurements, and within-person correlation of 0.5. MAF was 0.05 and proportion exposed at any given visit was 0.07. Outcome variance was higher in participants with any exposure, compared to participants with no exposure. P-values were estimated via GEE with an independence working correlation matrix, in the following contexts: (i) using sandwich standard error estimates and the standard normal reference distribution, (ii) using sandwich standard error estimates and a t reference distribution with Satterthwaite estimates of degrees of freedom, (iii) using Mancl and DeRouen’s alternate standard error estimates and a t reference distribution, (iv) Wang and Long’s alternate standard error estimates and a normal reference distribution, and (v) using a score test. Plots are quantile-quantile plots of −log10(pvalues), with the cone indicating the 95% prediction interval for the ordered −log10(p-values), when p-values are truly uniformly distributed between zero and one.



(a)

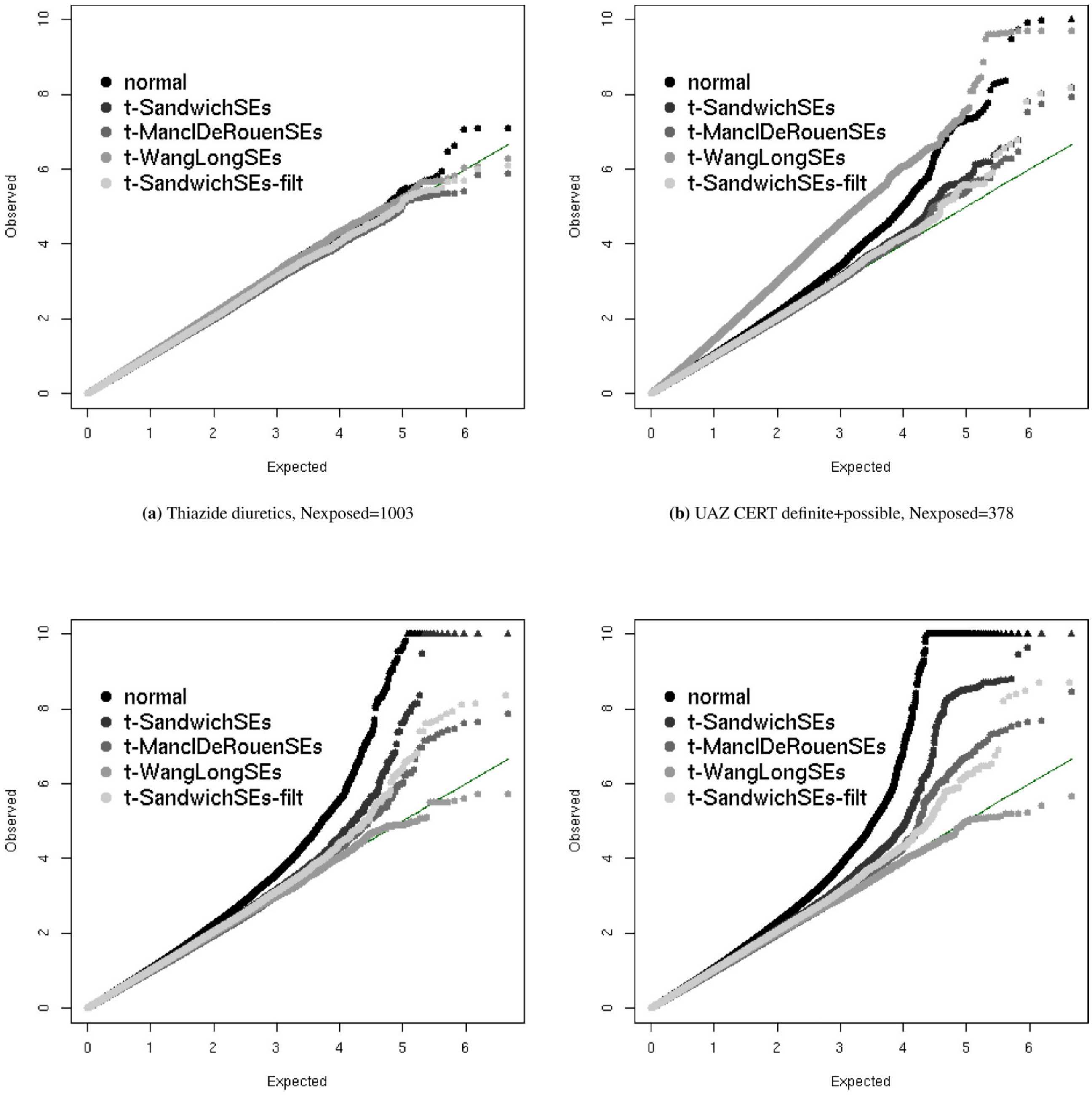
(b)

(c)

(d)

Figure 3.

Data from 1 million iterations were simulated under the null model, for cohort sizes of 1000, 1000, and 2000 individuals, each with 4 measurments, and within-person correlation of 0.5.We focused on analysis of gene-environment interactions, with MAF of 0.05 and probability of environmental exposure at any given visit of 0.05. Cohort-specific p-values (a,b,c) were estimated via GEE with an independence working correlation matrix, using either a normal reference distribution or a t reference distribution with degrees of freedom based on Satterthwaite's approximation. Meta-analytic p-values (d) were estimated using three different methods: 1) inverse-variance weighted z-statistics compared to a normal reference distribution, 2) inverse-variance weighted z-statistics compared to a t distribution with degrees of freedom equal to the sum of the individual cohort degrees of freedom, and 3) weighted z-value based meta-analysis of p-values using cohort-specific t reference distributions. Plots are quantile-quantile plots of −log10(pvalues), including only p-values less than 0.001.



(a) Thiazide diuretics, Nexposed=1003

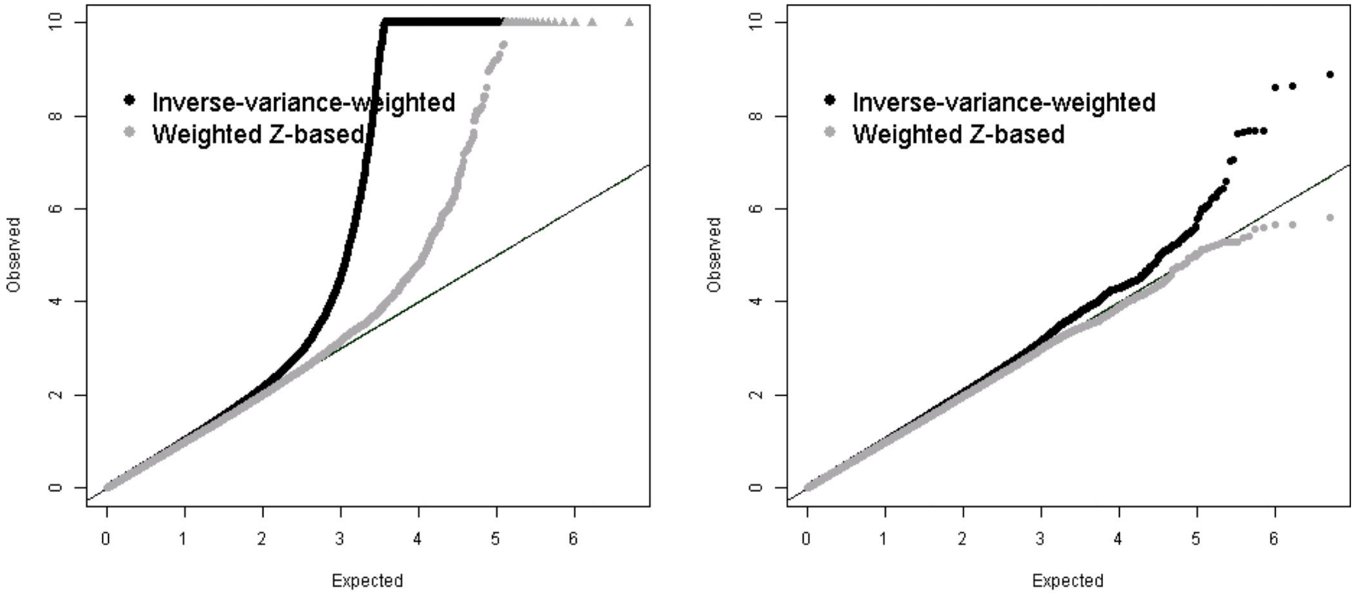
(b) UAZ CERT definite+possible, Nexposed=378

(c) Sulfonylureas, Nexposed=280

(d) Tri- and Tetra-cyclic Antidepressants, Nexposed=191

Figure 4.

CHS genome-wide quantile-quantile plots of −log10(pvalues) for gene-drug interactions, with three different standard error etimates and using either a normal reference distribution or a t reference distribution with Satterthwaite estimates of degrees of freedom, as specified. For sandwich variance estimates with a t reference distribution, the dark gray points represent all SNPs, whereas the lightest gray points exclude those SNPs with approxdf ≤ 10.



(a) All SNPs

(b) Excluding SNPs with approxdf ≤ 10

Figure 5.

Quantile-quantile plot of −log10(p-values) from meta-analysis of ARIC, CHS, RS1 and RS2 gene-UAZ CERT interaction estimates.